



Développement de modèles mécanistiques et évaluation de l'incertitude des paramètres par bootstrap : application aux médicaments anti-angiogéniques

Hoai-Thu Thai

► To cite this version:

Hoai-Thu Thai. Développement de modèles mécanistiques et évaluation de l'incertitude des paramètres par bootstrap : application aux médicaments anti-angiogéniques. Pharmacologie. Université Paris-Diderot - Paris VII, 2013. Français. NNT : . tel-00881823

HAL Id: tel-00881823

<https://theses.hal.science/tel-00881823>

Submitted on 9 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE : 393
SANTÉ PUBLIQUE : ÉPIDÉMIOLOGIE ET SCIENCES DE
L'INFORMATION BIOMÉDICALE

DOCTORAT

Hoai Thu THAI

DÉVELOPPEMENT DE MODÈLES MÉCANISTIQUES ET ÉVALUATION DE L'INCERTITUDE
DES PARAMÈTRES PAR BOOTSTRAP : APPLICATION AUX MÉDICAMENTS
ANTI-ANGIOGÉNIQUES

Thèse dirigée par le Dr Emmanuelle Comets

Soutenue le 24 mai 2013

JURY

Monsieur Jean-Yves Mary	Président du Jury
Monsieur Etienne Chatelut	Rapporteur
Monsieur Didier Concordet	Rapporteur
Monsieur Xavier Paoletti	Examineur
Madame Emmanuelle Comets	Directrice de thèse
Madame Christine Veyrat-Follet	Co-encadrante

*A ma famille,
A mes amours,*

Avec tous mes remerciements

A ma directrice de thèse, Emmanuelle Comets, pour sa grande patience et sa disponibilité. Grâce à son exceptionnelle réactivité face à mes nombreuses interrogations, j'ai pu bénéficier d'un encadrement "à distance" particulièrement efficace. Ses compétences scientifiques ainsi que ses qualités humaines m'ont permis d'achever cette thèse.

A Etienne Chatelut et Didier Concordet d'avoir accepté d'évaluer mon travail de thèse en tant que rapporteurs. Leurs commentaires m'ont aidée à améliorer ce manuscrit.

A Jean-Yves Marie de m'avoir fait l'honneur de présider mon jury de thèse et de m'avoir guidée dans les démarches administratives.

A Xavier Paoletti d'avoir accepté de participer à mon jury de thèse.

A France Mentré de m'avoir chaleureusement accueillie au sein de son laboratoire avec de bonnes conditions de travail. Sa compétence, sa rigueur scientifique et sa sagacité m'ont beaucoup appris.

A Christine Veyrat-Follet pour son encadrement de mon stage de master puis de ma thèse au sein du groupe de Modeling and Simulation chez Sanofi, qui a financé cette thèse. Ses expériences et ses encouragements m'ont été d'un grand soutien.

A Catherine Dubruc pour le projet de thèse qu'elle a élaboré et pour la confiance qu'elle m'a accordée.

A Nick Holford pour ses conseils et sa collaboration sur mes travaux de bootstrap.

A l'ensemble des collaborateurs et des ex-collaborateurs dans le département Drug Disposition de Sanofi avec une pensée particulière à Gerard Sanderink, François Donat, Paul Chevalier, Nathalie Fagniez pour les discussions scientifiques. Merci à Nicole, Jean-Baptiste, Claire pour leur aide précieuse.

A Houda, Anne, Quentin, Thu-Thuy, Cyrielle, Cédric, Hervé, Marie, Tram et à tous les autres membres de l'UMR 738 INSERM Paris Diderot pour m'avoir intégrée au sein de leur équipe et pour l'ambiance très agréable tout au long de ces trois années.

A Julie, Caroline, Steven, François, Rebeca et plus récemment Adrien, qui ont partagé le même bureau et m'ont supportée au quotidien.

A tous mes amis pour leur amitié, leur compréhension et leur soutien. Un grand merci à mes relecteurs Aude, Christine et Marie.

A toute ma famille pour leur soutien inconditionnel, à ma mère de m'avoir orientée dans les études de pharmacie, à mes beaux-parents pour leur aide précieuse particulièrement durant la rédaction de ma thèse.

A mon père qui n'a pas pu me voir grandir mais je sais qu'il est à mes côtés à jamais.

A Thien-Phu, mon cher mari pour m'avoir encouragée et supportée durant toute la thèse. Son amour, son humour et ses échanges scientifiques à la maison m'ont beaucoup aidée.

Enfin, à mon bout'chou Minh-Dang avec sa joie de vivre qui m'a aidé à ne pas baisser les bras dans la réalisation de ma thèse. Sa venue était un événement très heureux et a constitué un "chapitre" spécial, inséparable de la thèse. Qu'il trouve ici mon amour immense pour lui derrière les mots. Cette thèse lui est entièrement dédiée.

Table des matières

1	Introduction	10
1.1	Les thérapies ciblées anti-angiogéniques en cancérologie	10
1.1.1	L'angiogenèse physiologique et tumorale	11
1.1.2	Le rôle du facteur de croissance de l'endothélium vasculaire (VEGF) dans la croissance tumorale	14
1.1.3	Les agents anti-angiogéniques	15
1.2	La modélisation pharmacocinétique (PK) et pharmacodynamique (PD) . .	18
1.2.1	Les relations dose-concentration-effet	18
1.2.2	Les modèles PK	19
1.2.3	Les modèles PD	22
1.3	L'analyse PK/PD de population	23
1.3.1	Les modèles non linéaires à effets mixtes	23
1.3.2	Les méthodes d'estimation	24
1.3.3	Le calcul de l'incertitude des paramètres	25
1.4	L'approche bootstrap pour l'estimation de l'incertitude	25
1.4.1	Le principe du bootstrap	26
1.4.2	Les méthodes du bootstrap en régression linéaire	26
1.4.3	Les méthodes du bootstrap dans les modèles non linéaires à effets mixtes	27
1.5	Objectifs de la thèse	28
2	Développement des modèles mécanistiques pour la PK et la PD d'affibercept dans le traitement du cancer	30
2.1	Développement d'un modèle mécanistique pour la pharmacocinétique d'affibercept libre et lié chez les volontaires sains	30
2.1.1	Résumé	30
2.1.2	Article 1 (publié)	31
2.2	Analyse de pharmacocinétique de population d'affibercept chez les patients atteints de cancer	45
2.2.1	Résumé	45
2.2.2	Article 2 (accepté)	46

2.3	Modélisation des données longitudinales de réponse tumorale et de l'exposition à l'affibercept chez les patients atteints de cancer du côlon . .	79
2.3.1	Résumé	79
2.3.2	Rapport technique	80
3	L'approche bootstrap pour l'estimation d'incertitude des paramètres dans le cadre des modèles non linéaires à effets mixtes	98
3.1	Développement et comparaison de différentes approches bootstrap dans les modèles linéaires à effets mixtes	98
3.1.1	Résumé	98
3.1.2	Article 3 (publié)	99
3.2	Evaluation des approches bootstrap dans les modèles non linéaires à effets mixtes	120
3.2.1	Résumé	120
3.2.2	Article 4 (soumis)	121
3.2.3	Application du bootstrap aux données réelles	147
4	Conclusion générale	150

Productions scientifiques liées à la thèse :

Articles publiés

THAI H.T., VEYRAT-FOLLET C., VIVIER N., DUBRUC C., SANDERINK G., MENTRÉ F., COMETS E. A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects. *British Journal of Clinical Pharmacology*, 2011, vol.72(3), p.402-414.

THAI H.T., MENTRÉ F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E. A comparison of bootstrap approaches for estimating standard errors of parameters in linear mixed effects models. *Pharmaceutical Statistics*, 2013, doi :10.1002/pst.1561.

Article accepté

THAI H.T., VEYRAT-FOLLET C., MENTRÉ F., COMETS E. Population pharmacokinetic analysis of free and bound aflibercept in patients with advanced solid tumors. *Cancer Chemotherapy and Pharmacology*, 2013, accepted.

Article soumis

THAI H.T., MENTRÉ F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models : a simulation study in population pharmacokinetics. Soumission dans *Journal of Pharmacokinetics and Pharmacodynamics*.

Communications orales

THAI H.T., VEYRAT-FOLLET C., MENTRÉ F., COMETS E. Population pharmacokinetic analysis of free and bound aflibercept in patients with advanced solid tumors. *Journées de Groupe Métabolisme et Pharmacocinétique*, Paris, France, Octobre 2012.

THAI H.T., MENTRÉ F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E. Bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models. *22th Population Group Approach in Europe*, Glasgow, Ecosse, Juin 2013.

Posters

THAI H.T., VEYRAT-FOLLET C., VIVIER N., DUBRUC C., SANDERINK G., MENTRÉ F., COMETS E. A mechanism-based model for the population pharmacokinetics of aflibercept in healthy subjects. *6th International Symposium on Measurement & Kinetics of In Vivo Drug Effects*, Leiden, Pays-bas, Avril 2010.

THAI H.T., MENTRÉ F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E. A comparison of bootstrap approaches for estimating standard errors of parameters in linear

mixed effects models. *20th Population Group Approach in Europe*, Athènes, Grèce, Juin 2011.

THAI H.T., MENTRÉ F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E. Residual-based bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models with heteroscedastic error. *21th Population Group Approach in Europe*, Venise, Italie, Juin 2012.

THAI HT., VEYRAT-FOLLET C., MENTRÉ F., COMETS E. Population pharmacokinetic analysis of free and bound aflibercept in patients with advanced solid tumors. *World Conference on Pharmacometrics*, Séoul, Corée du Sud, Septembre 2012.

Chapitre 1

Introduction

Ce travail porte sur le développement des modèles mécanistiques pour les médicaments anti-angiogéniques qui inhibent le facteur de croissance de l'endothélium vasculaire (VEGF). Ces nouveaux médicaments sont en pleine émergence dans le traitement du cancer mais leurs propriétés pharmacocinétiques (PK) et pharmacodynamiques (PD) complexes (liées au VEGF) n'ont pas été bien caractérisées. Il est donc nécessaire de développer des modèles plus mécanistiques pour mieux comprendre la cinétique de la liaison médicament-récepteur et mieux évaluer l'efficacité du médicament. La construction de ces modèles demande de prendre en compte des données issues de plusieurs doses. Dans ce cadre, les modèles non linéaires à effets mixtes (MNLEM) fournissent un bon outil pour analyser un ensemble de sujets. Ils permettent l'estimation des paramètres typiques de population et de leur variabilité inter-sujet. L'incertitude liée à l'estimation des paramètres dans des modèles complexes peut être biaisée et parfois n'est pas obtenue. L'approche bootstrap fournit une technique alternative pour l'estimer mais elle n'a pas encore été évaluée de façon approfondie dans les MNLEM.

Dans la partie 1.1, nous introduisons de façon générale les thérapies ciblées anti-angiogéniques en cancérologie. Nous décrivons ensuite brièvement la modélisation PK et PD dans la partie 1.2 et présentons l'approche de population pour analyser les données PK et PD dans la partie 1.3. La partie 1.4 introduit les méthodes du bootstrap dans les modèles à effets mixtes. Enfin, nous présentons les objectifs de la thèse dans la partie 1.5.

1.1 Les thérapies ciblées anti-angiogéniques en cancérologie

Le cancer touche environ 13 millions de personnes dans le monde entier en 2010 et est aujourd'hui la première cause de mortalité chez les hommes et la deuxième chez les femmes après les maladies cardiovasculaires en France. Il y a 357 700 nouveaux cas de cancers estimés en 2010. Environ 150 000 personnes meurent chaque année d'un cancer en France soit plus de 400 décès par jour. Chez l'homme, le cancer du poumon est la principale cause de décès, suivi par les cancers colorectal et de la prostate. Par contre, chez la femme, le cancer du sein est la principale cause de décès suivi par les cancers

colorectal et du poumon. Le traitement du cancer repose principalement sur la chirurgie, la radiothérapie et la chimiothérapie. Cependant, la lutte contre le cancer est loin d'être achevée avec les traitements anticancéreux standard présentant des effets secondaires parfois sévères. Le développement de molécules plus spécifiques est donc nécessaire pour améliorer l'efficacité et réduire la toxicité des traitements. Dans le cadre de cette thèse, nous nous sommes intéressés aux nouvelles thérapies anti-angiogéniques pour le traitement des tumeurs solides.

1.1.1 L'angiogenèse physiologique et tumorale

L'angiogenèse est le processus de formation de nouveaux vaisseaux sanguins (néovascularisation) à partir de vaisseaux préexistants. Ce phénomène est indispensable au cours de nombreux processus physiologiques tels que le développement embryonnaire, l'implantation du placenta, mais aussi pathologiques tels que la croissance des tumeurs malignes ou le développement des métastases.

Angiogenèse physiologique

Il s'agit d'un constituant fondamental du développement embryonnaire qui est le résultat de deux processus distincts : la vasculogenèse et l'angiogenèse. Lors de la vasculogenèse, les cellules précurseurs indifférenciées (angioblastes) chez l'embryon se différencient en cellules endothéliales, qui prolifèrent et s'assemblent pour donner un réseau de vaisseaux primitifs et peu fonctionnels. Ensuite, lors de l'angiogenèse, ce dernier est remodelé en un réseau vasculaire mature comprenant les artères, les veines et les capillaires (CARMELIET, 2000; COUFFINHAL et al., 2001). Le terme "angiogenèse" a été introduit pour la première fois par Hunter en 1787.

Chez l'adulte, le réseau vasculaire est quiescent. Cependant, l'angiogenèse s'observe physiologiquement dans certaines conditions, par exemple lors du cycle menstruel féminin, pendant la gestation au niveau de la glande mammaire et du placenta, au cours de l'inflammation, des processus de réparation tissulaire et lors d'adaptation des tissus à une hypoxie.

Trois mécanismes sont à l'origine de la formation de nouveaux vaisseaux sanguins (CONWAY et al., 2001) : le bourgeonnement des vaisseaux préexistants vers des stimuli angiogéniques (*sprouting*), la division en vaisseaux secondaires par invagination des parois ou l'intussusception (*splitting*), et la division des vaisseaux préexistants par insertion de colonnes interstitielles dans leur lumière (*bridging*). L'angiogenèse par bourgeonnement (Figure 1.1) est le mécanisme le mieux défini. Dans ce processus, des cellules endothéliales (CE) sont d'abord activées par les facteurs de croissance liés à leurs récepteurs, ce qui conduit à la dégradation de la matrice extracellulaire (MEC) et de la membrane basale entourant les CE. Cela permet aux CE d'envahir la matrice environnante et, par la suite,

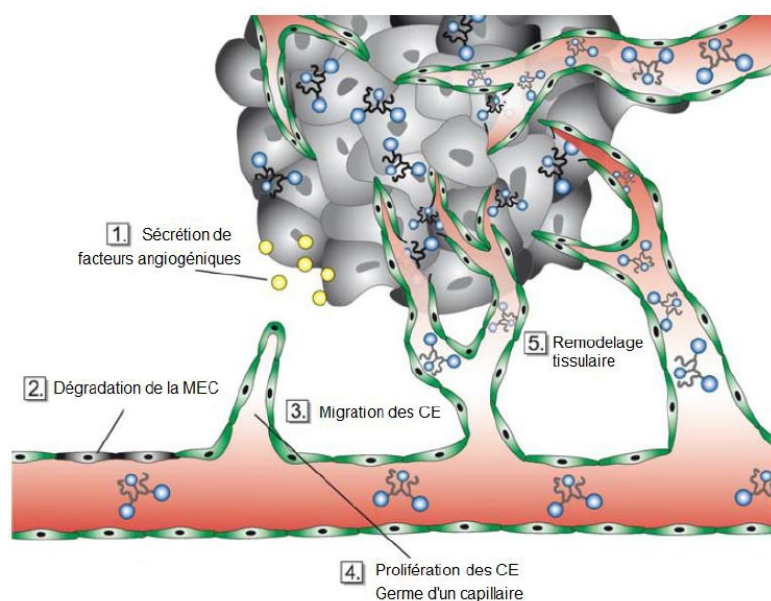


FIGURE 1.1 – Etapes principales de l'angiogenèse par bourgeonnement
Source : SATCHI-FAINARO et al. (2006)

de proliférer et de migrer à travers la matrice. Enfin, par polarisation, les CE s'organisent en structures tubulaires avec une nouvelle lame basale, des germes (*sprouts*) se connectent aux vaisseaux voisins, et forment un vaisseau sanguin immature. La stabilisation des ces vaisseaux immatures est établie par le recrutement de cellules murales et la mise en place de la MEC (HILLEN et GRIFFIOEN, 2007).

L'angiogenèse n'est pas contrôlée par un seul facteur mais par un équilibre entre facteurs pro-angiogéniques (activateurs) et facteurs anti-angiogéniques (inhibiteurs). Les activateurs angiogéniques sont les enzymes protéolytiques qui interviennent dans la dégradation de la membrane basale telles que les métalloprotéases matricielles (MMPs), les facteurs de croissance qui stimulent la migration et la prolifération des CE tels que le VEGF (*Vascular Endothelial Growth Factor*) qui est le facteur le plus important, l'EGF (*Epidermal Growth Factor*), le PDGF (*Platelet Derived Growth Factor*), l'IGF (*Insulin Growth Factor*), le TNF (*Transforming Growth Factor*), le FGF (*Fibroblast Growth Factor*) et les molécules d'adhésion cellulaire telles que les intégrines, E-sélectine. Les inhibiteurs angiogéniques sont les inhibiteurs des enzymes protéolytiques tels que l' α 2M (α -2-macroglobuline), les TIMPs (*Tissue Inhibitors of Metalloproteinases*), le TFPI-2 (*Tissue Factor Pathway Inhibitor-2*) et les inhibiteurs de l'activation des CE tels que le TSP-1 (thrombospondine-1), l'angiostatine, l'endostatine, le PF-4 (*Platelet Factor 4*), l'IFN- α (*Interferon- α*).

Angiogenèse tumorale

Le rôle crucial de l'angiogenèse dans la croissance tumorale a été proposé pour la première fois par Folkman en 1971. L'angiogenèse tumorale correspond à l'apparition

de nouveaux vaisseaux destinés à répondre aux besoins métaboliques (oxygène et nutriments) des cellules tumorales et contribue également à leur survie (FOLKMAN, 1971). En effet, la croissance d'une tumeur solide est fortement dépendante de ce processus (FOLKMAN, 1972). Contrairement à l'angiogenèse physiologique, l'angiogenèse tumorale est désorganisée puisqu'elle est constituée de néovaisseaux de constitution anormale (dilatés, tortueux, et borgnes), immatures (absence de péricytes), responsables d'une densité vasculaire hétérogène.

L'étape la plus importante du phénomène d'angiogenèse tumorale est le "switch angiogénique", le passage de l'état "dormant" à une lésion vascularisée, caractérisé par un déséquilibre entre les facteurs pro-angiogéniques et les facteurs anti-angiogéniques. Cette angiogenèse résulte de l'interaction entre le cancer et le stroma lorsque la tumeur atteint la taille critique de $1-2 \text{ mm}^3$. Au delà de cette taille, une tumeur solide ne peut croître sans développer des nouveaux vaisseaux sanguins pour éviter l'hypoxie et l'apoptose des cellules tumorales.

Dans le cas des tumeurs solides, l'angiogenèse fait intervenir tous les acteurs de l'angiogenèse physiologiques et aussi des acteurs spécifiques comme les fibroblastes associés au cancer (CAFs) et les précurseurs endothéliaux circulants issus de la moelle osseuse (*bone-marrow-derived angiogenic cells*, BMC). Les principaux facteurs dans les tumeurs solides sont présentés dans la Figure 1.2 (FERRARA et KERBEL, 2005).

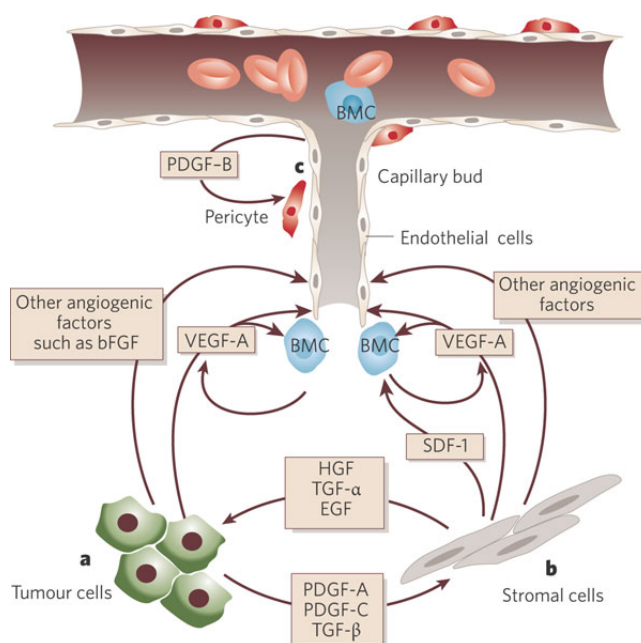


FIGURE 1.2 – Les principaux facteurs dans les tumeurs solides

Source : FERRARA et KERBEL (2005)

Les cellules tumorales (a) libèrent le VEGF-A et d'autres agents pro-angiogéniques qui stimulent la prolifération et la migration des CE adjacentes. Les cellules tumorales libèrent également les facteurs de recrutement des cellules stromales comme le PDGF-A,

le PDGF-C ou le TGF- β . Les cellules du stroma (b) constituent une source supplémentaire de facteurs pro-angiogéniques, y compris le VEGF-A et le SDF-1 (*Stromal Cell-derived Factor-1*) qui induisent le recrutement des BMC. Les CE (c) libèrent le PDGF- β permettant de recruter des péricytes dans la micro-vascularisation.

1.1.2 Le rôle du facteur de croissance de l'endothélium vasculaire (VEGF) dans la croissance tumorale

La famille du VEGF comprend 6 facteurs de croissance impliqués dans l'angiogenèse : le VEGF-A dont il existe 6 isoformes, le VEGF-B, le VEGF-C, le VEGF-D, le VEGF-E et le PlGF (*Placenta Growth Factor*). Parmi ces facteurs, le VEGF-A est le facteur le plus important. Il est surexprimé dans la plupart des tumeurs solides et son rôle clé dans l'angiogenèse tumorale a été démontré dans plusieurs études (FERRARA, 2007). L'inactivation d'un seul allèle du VEGF-A est associée à une mortalité embryonnaire provoquée par une vascularisation défectueuse (FERRARA, 2004; YANCOPOULOS et al., 2000).

Le rôle du VEGF dans la croissance tumorale est illustré dans la Figure 1.3.

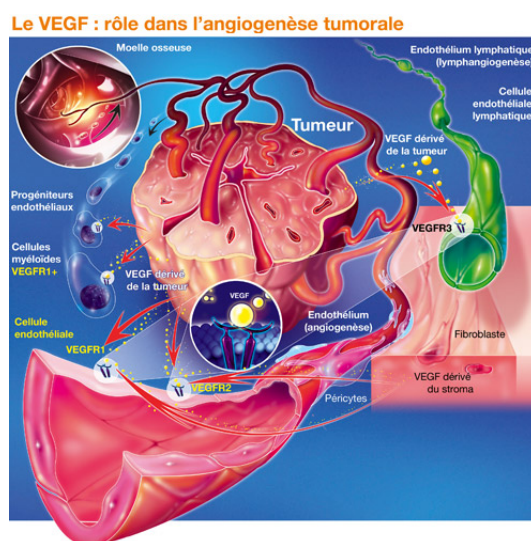


FIGURE 1.3 – Le rôle du VEGF dans la croissance tumorale

Source : <http://www.illustration-medicales.fr/>

Pour exercer ses propriétés angiogéniques, le VEGF se lie principalement à trois récepteurs à activité tyrosine kinase : le VEGFR-1 dont le rôle précis reste inconnu et le VEGFR-2 qui est impliqué dans la majorité des effets angiogéniques, le VEGFR-3 qui joue un rôle uniquement dans la lymphangiogenèse. Le VEGFR-1 et le VEGFR-2 sont essentiellement exprimés au niveau des cellules endothéliales vasculaires, alors que le VEGFR-3 est localisé principalement au niveau des cellules endothéliales lymphatiques. La fixation du VEGF sur la partie extracellulaire de son récepteur induit la dimérisation et

l'activation de l'activité tyrosine kinase intracellulaire, conduisant à la prolifération et à la survie cellulaire. Cette activation des CE conduit aussi à la sécrétion de différents facteurs et enzymes qui sont très utiles à la dégradation de la matrice extracellulaire indispensable à la prolifération des tubes endothéliaux (HICKLIN et ELLIS, 2005). L'ensemble de ces actions biologiques favorise l'angiogenèse et la formation d'un microenvironnement favorable à la croissance des cellules tumorales.

1.1.3 Les agents anti-angiogéniques

La découverte du VEGF dans l'angiogenèse tumorale en 1989 (LEUNG et al., 1989) a conduit au développement d'une nouvelle approche pour le traitement du cancer dite la thérapie anti-angiogénique. Cette thérapie a attiré l'intérêt de nombreux chercheurs grâce à des avantages théoriques : (i) elle agit sur les CE de l'environnement péri-tumoral et non pas sur la cellule tumorale elle-même comme les agents cytotoxiques, (ii) son spectre d'activité est potentiellement large car l'angiogenèse est un processus rencontré dans la plupart des tumeurs solides, (iii) sa bonne tolérance car l'angiogenèse physiologique est un phénomène limité chez l'adulte. A ce jour, l'approche anti-angiogénique la plus prometteuse est celle ciblant la cascade du VEGF/VEGFR (HOLASH et al., 2002). Il existe plusieurs voies d'inhibition de cette cascade : inhibition du ligand (VEGF) (e.g les anticorps anti-VEGF ou les récepteurs du VEGF soluble), inhibition de ses récepteurs (e.g les anticorps anti-VEGFR-1 et anti-VEGFR-2) et inhibition du domaine tyrosine kinase des récepteurs (e.g les inhibiteurs de tyrosine kinase) (Figure 1.4).

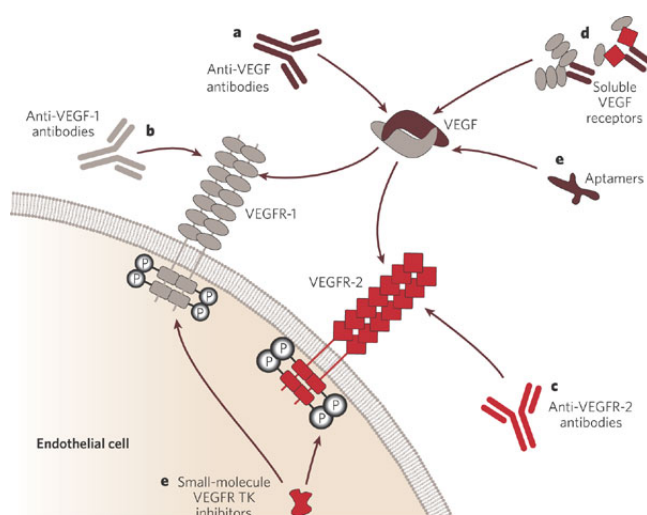


FIGURE 1.4 – Les stratégies d'inhibition de la cascade du VEGF
Source : FERRARA et KERBEL (2005)

Au cours de la dernière décennie, la FDA (*Food and Drug Administration*, Etats-Unis) a approuvé huit agents anti-angiogéniques pour le traitement du cancer dont le premier

a été le bévacizumab, un anticorps anti-VEGF qui bloque le VEGF-A, en 2004 (Table 1.1). Un grand nombre d'autres agents anti-angiogéniques sont actuellement en phase III de développement clinique. Tous les médicaments anti-angiogéniques approuvés sont les agents anti-VEGF. Le bévacizumab et l'aflibercept bloquent le ligand alors que d'autres molécules sont des petites molécules qui agissent par inhibition de l'activité de la tyrosine kinase des récepteurs du VEGF.

TABLE 1.1 – Les agents anti-angiogéniques approuvés par la FDA pour le traitement du cancer (2004-2012)

Source : DUDA (2012)

Anti-VEGF drug	Mecanism of action	Approved indication	Company
Bevacizumab (Avastin)	Blocks VEGF-A	Metastatic colorectal cancer (with chemotherapy) Metastatic nonsquamous non-small cell lung cancer (with chemotherapy) Metastatic breast cancer (with chemotherapy) Recurrent glioblastoma (monotherapy) Metastatic renal cell carcinoma (with IFN α)	Genetech/Roche (Basel/Switzerland)
Sunitinib (Sutent)	Inhibits receptor tyrosine kinase signaling	Metastatic renal cell carcinoma (monotherapy) Gastrointestinal stromal tumors (monotherapy) Pancreatic neuroendocrine tumors (monotherapy)	Pfizer (New York, NY)
Sorafenib (Nexavar)	Inhibits receptor tyrosine kinase signaling	Metastatic renal cell carcinoma Unresectable hepatocellular carcinoma	Bayer/Onyx (South San Francisco, CA)
Pazopanib (Votrient)	Inhibits receptor tyrosine kinase signaling	Metastatic renal cell carcinoma Advanced soft tissue sarcoma	GlaxoSmithKline (London, UK)
Vandetanib (Zactima)	Inhibits receptor tyrosine kinase signaling	Advanced medullary thyroid cancer	AstraZeneca (London, UK)
Axitinib (Inlyta)	Inhibits receptor tyrosine kinase signaling	Advanced renal cell carcinoma	Pfizer (New York, NY)
Regorafenib (Stivarga)	Inhibits receptor tyrosine kinase signaling	Metastatic colorectal cancer	Bayer/Onyx (South San Francisco, CA)
Aflibercept (Zaltrap)	Binds VEGF-A, VEGF-B and PlGF	Metastatic colorectal cancer (with chemotherapy)	Regeneron/Sanofi (Paris)

La thérapie anti-angiogénique permet de stabiliser la progression de la tumeur, comme cela a été démontrée dans plusieurs cancers, conduisant à l'amélioration de la survie sans progression et la survie globale comparé au traitement standard (RAPISARDA et MELILLO, 2012; JAIN et al., 2009). Malgré ce premier succès, il existe quelques limitations de l'application de cette stratégie en oncologie. Tout d'abord, ce traitement n'a pas montré le même avantage dans d'autres cancers avancés par exemple le cancer du sein, le cancer de l'ovaire, le cancer de la prostate et le cancer pancréatique. Deuxièmement, certains patients répondent très peu ou pas du tout au traitement anti-angiogénique. Ces résultats peuvent être liés au développement de résistances extrinsèques et intrinsèques à la thérapie anti-angiogénique (BERGERS et HANAHAN, 2008; KIERAN et al., 2012). Un des mécanismes de la résistance extrinsèque est la stimulation d'une cascade alternative de l'angiogenèse, par exemple celle du facteur de croissance placentaire (PlGF) quand la cascade principale du VEGF-A est bloquée. A ce jour, il n'existe pas encore de biomarqueurs validés permettant de surveiller l'efficacité, d'optimiser la posologie, de prédire la toxicité ou la résistance de

la thérapie anti-VEGF et de sélectionner les patient répondeurs (JAIN et al., 2009; LOGES et al., 2010). Cependant, un grand nombre de biomarqueurs systémiques, circulants, tissulaires et d'imagerie ont été proposés et nécessitent des validations prospectives.

Dans cette thèse, nous avons étudié le mécanisme d'action d'un nouvel anti-angiogénique, l'affibercept (également appelé ziv-affibercept aux États-Unis, développé conjointement par Sanofi et Regeneron Pharmaceuticals) (CHU, 2009). Il s'agit une protéine de fusion comportant des domaines des récepteurs VEGFR-1 et VEGFR-2 et une portion Fc de l'immunoglobuline G1 humaine (IgG1) (Figure 1.5).

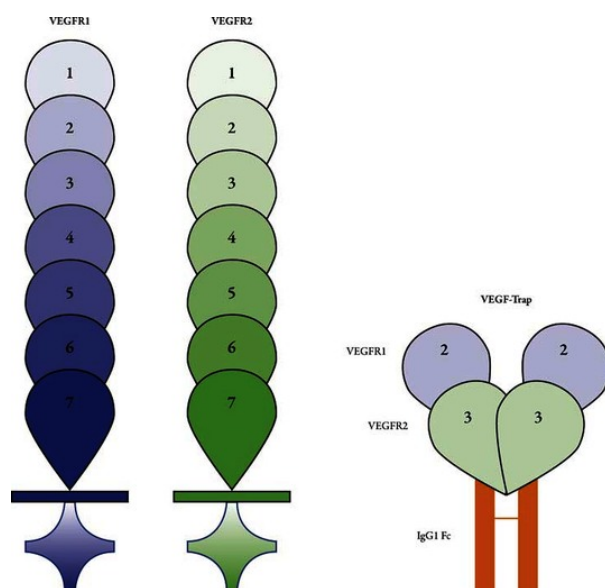


FIGURE 1.5 – Structure moléculaire de l'affibercept (VEGF-Trap)
Source : SOPHIE et al. (2012)

L'affibercept se fixe au VEGF pour former un complexe (VEGF :affibercept) dans le sang et dans l'espace extravasculaire et empêche le VEGF d'interagir avec ses récepteurs sur les CE. Contrairement au bécacizumab, qui se lie uniquement au VEGF-A et forme des complexes mutlimériques, l'affibercept se lie à différentes formes du VEGF-A, mais également au VEGF-B et au facteur de croissance placentaire (PlGF). L'affinité de l'affibercept pour VEGF-A (K_d *in vitro* = 0.5 pM) est beaucoup plus forte que celle du bécacizumab (K_d *in vitro* = 500 pM). En ciblant à la fois le VEGF-A et le PlGF, l'affibercept permet probablement de réduire le développement de la résistance aux thérapies anti-VEGF (GAYA et TSE, 2012).

Les études précliniques chez la souris ont démontré les effets de l'affibercept sur la régression capillaire, le rétrécissement des vaisseaux, la cessation du flux sanguin ainsi que l'apoptose des CE (BAFFERT et al., 2006). L'inhibition de la croissance tumorale, l'inhibition des métastases et l'amélioration de la survie ont été observées avec l'affibercept dans les xénogreffes de tumeurs de différents cancers (HOLASH et al., 2002; EICHEN et al., 2012). Les propriétés PK ont été évaluées tout d'abord dans les études de phase

I chez les volontaires sains après une administration par voie intraveineuse. Selon les résultats de l'analyse non compartimentale, la PK de l'affibercept est caractérisée par une clairance dose dépendante (allant de 1,22 L/jour pour la dose de 1 mg/kg à 0,78 L/jour pour la dose de 4 mg/kg). Le volume de distribution à l'état d'équilibre (V_{ss}) était environ de 6 L. L'activité antitumorale de l'affibercept a été étudiée dans des études de phase I à doses croissantes, de phase II et de phase III, en monothérapie ou en association avec des chimiothérapies chez les patients atteints du cancer du poumon, de l'ovaire, du pancréas et du cancer colorectal (GAYA et TSE, 2012). Récemment, il a été montré que l'administration de l'affibercept en perfusion intraveineuse à 4 mg/kg toutes les 2 semaines en combinaison avec le FOLFIRI (5-fluorouracil (5-FU), la leucovorine et l'irinotécan), améliore la survie globale, la survie sans progression et le taux de réponse chez les patients atteints du cancer colorectal métastatique préalablement traités avec l'oxaliplatine (VAN CUTSEM et al., 2012). Cette indication a récemment été approuvée comme traitement de deuxième ligne par les agences réglementaires aux États-Unis et en Europe.

1.2 La modélisation pharmacocinétique (PK) et pharmacodynamique (PD)

1.2.1 Les relations dose-concentration-effet

Comme précédemment expliqué dans la partie 1.1, les médicaments anti-angiogéniques offrent une nouvelle thérapie plus efficace pour le traitement du cancer. Possédant une structure originale, l'affibercept a une plus forte affinité pour le VEGF-A que d'autres anticorps monoclonaux et se lie également au VEGF-B et au PlGF. Dans le cadre de cette thèse, nous avons cherché à mieux comprendre les interactions de l'affibercept avec l'organisme, autrement dit les relations entre la dose du produit administré, les concentrations mesurées et l'effet du médicament (Figure 1.6). La relation entre la dose et la concentration dans l'organisme est caractérisée par les études pharmacocinétiques (PK) alors que la relation entre la concentration et son effet sur l'organisme est caractérisée par les études pharmacodynamiques (PD).

Dans le développement des médicaments, la modélisation PK/PD joue un rôle de plus en plus important pour mieux étudier les relations dose-concentration-effet (BONATE, 2011; AARONS et al., 2001; LALONDE et al., 2007). Les connaissances acquises aux précédentes étapes du développement sont intégrées dans un modèle structural régit par un nombre limité de paramètres qui prennent en compte les processus physiologiques connus. Une bonne paramétrisation du modèle structural permet de quantifier des paramètres d'intérêt physiologique et de mieux utiliser le modèle en mode explicatif (voire dans

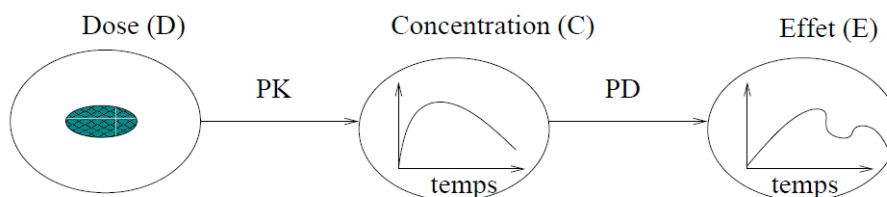


FIGURE 1.6 – Relations dose-concentration-effet

certaines situations en prédictif).

1.2.2 Les modèles PK

La pharmacocinétique étudie en fonction du temps, les différents processus d'absorption, de distribution, de métabolisme et d'élimination (ADME) qui gouvernent le devenir du médicament dans l'organisme auquel il a été administré (GABRIELSSON et WEINER, 1999). La modélisation PK classique consiste à considérer l'organisme comme un ensemble de compartiments entre lesquels le médicament peut circuler. A titre d'exemple, la figure 1.7 représente un modèle à un seul compartiment, schématisant le devenir d'une dose de médicament administrée par voie orale, absorbée depuis l'intestin avec une constante d'absorption k_a vers un unique compartiment de volume V , et éliminée avec une constante d'élimination k_{el} .

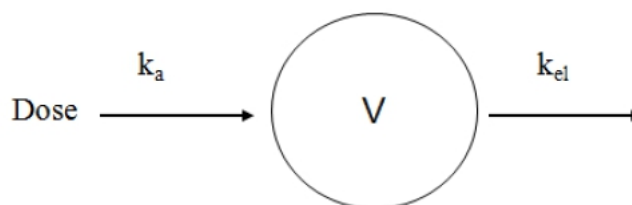


FIGURE 1.7 – Schéma d'un modèle à un compartiment, avec absorption et élimination d'ordre 1

Ce schéma permet de représenter l'évolution de la quantité du médicament au cours du temps $A(t)$ sous forme d'une équation différentielle qui dépend de la fonction administration du médicament $e(t)$, soit pour le modèle décrit par le schéma 1.7.

$$\frac{dA(t)}{dt} = e(t) - k_{el} A(t) \quad (1.1)$$

Une solution analytique de ce système, si elle existe, peut être utilisée pour formuler le modèle en une fonction mathématique du temps (GIBALDI et PERRIER, 1982) non

linéaire en ses paramètres PK. Par exemple, l'équation analytique du modèle décrit par le schéma 1.7 est écrite comme ci-dessous :

$$C(t) = \frac{D}{V} \frac{k_a}{k_a - k_{el}} (e^{-k_{el} t} - e^{-k_a t}) \quad (1.2)$$

La figure 1.8 présente les profils individuels des concentrations plasmatiques observées au cours du temps ainsi que le profil "moyen" correspondant pour un médicament administré par voie orale.

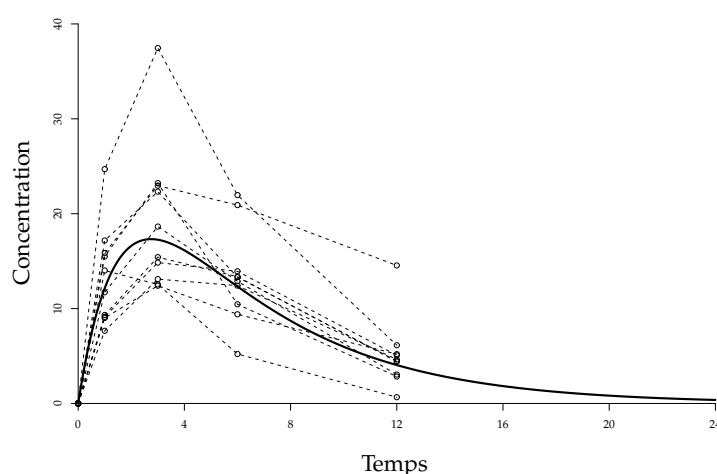


FIGURE 1.8 – Profil "moyen" des concentrations plasmatique (ligne continue) sous un modèle pharmacocinétique à un compartiment d'absorption et d'élimination d'ordre 1, superposé à différents profils individuels (lignes pointillés), après une dose unique.

Les modèles de disposition liés à une cible thérapeutique

Dans le cadre d'une cinétique linéaire, les constantes de vitesse (élimination, absorption, distribution) sont indépendantes des doses administrées tout comme les clairances. C'est-à-dire que les concentrations augmentent proportionnellement à la dose. Cependant, cette cinétique ne correspond pas aux caractéristiques des produits biologiques ayant une forte affinité avec la cible thérapeutique, y compris les médicaments anti-angiogéniques ciblant le VEGF. En effet, la PK de ces agents biologiques devient plus ou moins non linéaire avec une diminution dose-dépendante du volume de distribution ou de la clairance car elle est significativement affectée par la liaison médicament-cible (MAGER et JUSKO, 2001). Ce phénomène a été baptisé "*target-mediated drug disposition*" (TMDD), terme introduit pour la première fois par LEVY (1994). Un exemple classique de TMDD observé sur les données PK chez l'homme est celui des inhibiteurs de l'enzyme de conversion de l'angiotensine. MAGER et JUSKO (2001) ont par la suite développé un modèle général pour les médicaments ayant une pharmacocinétique de type TMDD.

Ce modèle décrit la voie d'élimination des concentrations plasmatiques du médicament comme la combinaison d'une élimination d'ordre un depuis le compartiment central et d'une élimination spécifique liée à la cible. En plus de la pharmacocinétique du médicament libre (i.e. médicament non lié à son récepteur cible), ce modèle décrit la production et la dégradation du récepteur, l'association et la dissociation du complexe médicament-cible (DR) et l'internalisation du complexe DR (Figure 1.9).

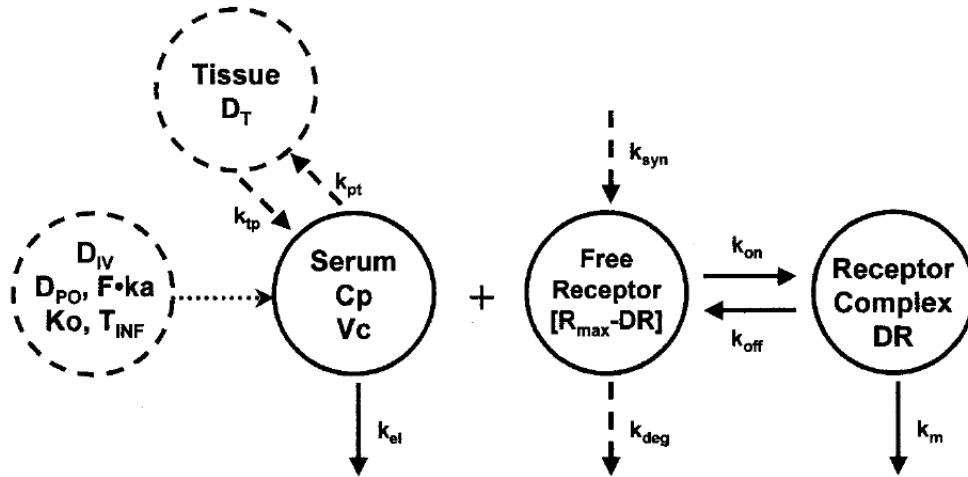


FIGURE 1.9 – Schéma du modèle TMDD général. Source : MAGER et JUSKO (2001)

Le médicament dans le compartiment central peut se distribuer dans des sites tissulaires non spécifiques (D_T), s'éliminer du système (k_{el}), ou se lier (k_{on}) à une cible pharmacologique pour former un complexe médicament-cible (DR). Ce complexe peut alors soit se dissocier (k_{off}) soit être internalisé et dégradé (k_m). La cible sous forme libre est contrôlée par sa synthèse (k_{syn}) et sa dégradation (k_{deg}).

Ce modèle implique plusieurs hypothèses :

- La liaison médicament-cible est un processus de liaison un-à-un simple (non coopérative ou allostérique)
- La liaison médicament-cible ne se produit que dans le compartiment central
- L'influence de la réponse immunitaire (par exemple, l'apparition de liaison et/ou neutralisation des anticorps) est négligeable
- Les vitesses de production et de dégradation de la cible sont constantes et elles ne dépendent pas des concentrations du médicament ou de la cible.

A ce jour, le TMDD s'est révélé le plus pertinent pour expliquer la cinétique des médicaments biologiques, tels que les anticorps monoclonaux, les cytokines, les facteurs de croissance, les protéines de fusion, les anticorps à petite molécule conjuguée, les hormones et des facteurs métaboliques (GIBIANSKY et GIBIANSKY, 2009). Le modèle TMDD central est appliqué à plusieurs produits d'origine biologique en modélisant les concentrations

moyennes du médicament libre, tels que le facteur inhibiteur de la leucémie humaine recombinante (rh-LIF) (SEGRAVE et al., 2004), l'interféron- β 1a recombinant (MAGER et al., 2003), et l'érythropoïétine recombinante (rHuEPO) (WOO et al., 2007). Il a également été étendu à un modèle TMDD périphérique MAGER (2006) et à un modèle plus complexe où le médicament se lie à plusieurs cibles (GIBIANSKY et GIBIANSKY, 2010a).

Le modèle TMDD complet est toutefois complexe, souvent surparamétré par rapport aux données disponibles. C'est la raison pour laquelle l'utilisation du modèle TMDD complet dans la modélisation PK/PD de population reste assez limitée (GIBIANSKY et GIBIANSKY, 2009). En conséquence, plusieurs approximations du modèle TMDD ont été proposées : QE (*quasi equilibrium*), QSS (*quasi steady state*) et MM (*Michaelis-Menten*) (MAGER et KRZYZANSKI, 2005; GIBIANSKY et al., 2008; GIBIANSKY et GIBIANSKY, 2010b). L'approximation QE est basée sur l'hypothèse que la liaison médicament-cible est beaucoup plus rapide que tous les autres processus du système. L'approximation QSS suppose que le changement des concentrations du complexe est plus lent que le processus de liaison et l'internalisation. L'approximation MM décrit le système lorsque la concentration de la cible est petite par rapport à la concentration du médicament libre, ce qui entraîne une saturation complète de la cible. Plus récemment, Gibiansky a proposé une deuxième approximation, dérivée de l'approximation MM, appelée IB-MM (*Irreversible Binding-MM*) pouvant être utilisée lorsque la constante de vitesse de dissociation est négligeable (GIBIANSKY et GIBIANSKY, 2010b).

1.2.3 Les modèles PD

La pharmacodynamie étudie les effets biochimiques et physiologiques du médicament sur l'organisme. L'activité biologique de la majorité des médicaments résulte de leurs interactions avec des récepteurs. Le modèle PD le plus simple est le modèle dit Emax, postulant un lien direct entre la concentration plasmatique et l'effet. Il consiste à modéliser les réponses instantanées comme la pression artérielle ou le débit cardiaque. Pour des mesures ayant une dynamique plus complexe, les modèles PD deviennent de plus en plus physiologiques et mécanistiques en reliant les processus biologiques, par exemple les modèles de réponse indirecte (JUSKO et KO, 1994) et les modèles de tolérance (GABRIELSSON et WEINER, 1999).

Dans le cas des médicaments anti-cancéreux, les modèles PD caractérisant l'évolution de taille de la tumeur au cours du temps sont de plus en plus utilisés pour évaluer l'efficacité des médicaments. Ce sont des modèles d'inhibition de la croissance tumorale ("*Tumor growth inhibition*", TGI) qui permettent de modéliser la dynamique des tumeurs. Plusieurs modèles TGI ont été développés dans la littérature. Le modèle empirique le plus simple comprend une décroissance exponentielle et une croissance linéaire (WANG et al.,

2009). Les modèles plus complexes décrivent la dynamique tumorale sous l'exposition du médicament et ainsi la résistance des tumeurs au traitement (THAM et al., 2008; CLARET et al., 2009; RIBBA et al., 2012).

1.3 L'analyse PK/PD de population

1.3.1 Les modèles non linéaires à effets mixtes

L'estimation des paramètres des modèles PK/PD décrits dans la section précédente peut être réalisée par plusieurs approches. Une première approche basée sur la régression non linéaire classique consiste à estimer les paramètres pour chaque sujet et puis calculer la moyenne et la variance de l'ensemble des paramètres individuels estimés (appelée méthode en deux étapes). Cette méthode requiert un nombre suffisant de données par sujet. Quand le modèle devient plus complexe, par exemple le modèle TMDD ou le modèle de réponse indirecte, le mécanisme sous-jacent du médicament ne peut pas être caractérisé par les données d'une seule dose. Il demande d'ajuster simultanément les données issues de plusieurs doses, y compris la forte dose qui sature l'effet du médicament. Dans ce cas, les mêmes paramètres peuvent être estimés pour tous les sujets en supposant une faible variabilité inter-sujet. Afin de pouvoir modéliser de manière plus pertinente les données de l'ensemble des sujets, notamment quand peu de données sont disponibles pour chaque sujet dans les essais cliniques, une approche plus complexe utilisant les modèles non linéaires à effets mixtes (MNLEM) peut être effectuée.

Cette approche, appelée approche de population, permet d'estimer les paramètres "moyens" de population ainsi que leur variabilité inter-sujet. Elle permet aussi de quantifier l'influence des covariables individuelles sur les différents paramètres PK/PD. Aujourd'hui, l'analyse de population fait l'objet de recommandations de la part des agences du médicament (EMEA, 1996; FOOD et ADMINISTRATION, 1999) et fait partie intégrante du dossier d'autorisation pour un nouveau médicament.

Dans cette thèse, nous avons utilisé cette méthodologie, que nous allons maintenant détailler, en commençant par les modèles statistiques. Nous noterons N le nombre de sujets dans l'étude. L'observation y_{ij} du sujet i ($i = 1, \dots, N$) au temps de prélèvement j ($j = 1, \dots, n_i$) est décrite par une fonction non linéaire f dépendant du temps t_{ij} et du vecteur des paramètres individuels ϕ_i du sujet i :

$$y_{ij} = f(t_{ij}, \phi_i) + g(t_{ij}, \phi_i)\epsilon_{ij} \quad (1.3)$$

Comme la fonction f qui correspond au modèle structural choisi, la fonction g est une fonction non linéaire des paramètres individuels. Elle représente la forme du modèle

d'erreur qui peut être un modèle constant $g(t_{ij}, \phi_i) = a$, proportionnel $g(t_{ij}, \phi_i, \xi) = bf(t_{ij}, \phi_i)$ ou encore combiné $g(t_{ij}, \phi_i) = a + bf(t_{ij}, \phi_i)$. Les erreurs résiduelles ϵ_{ij} sont supposées indépendantes et identiquement distribuées selon une distribution gaussienne de moyenne nulle et de variance 1 ($\epsilon_{ij} \sim \mathcal{N}(0, 1)$). Le vecteur de paramètres ϕ_i se décompose en un vecteur d'effets fixes μ , une matrice de covariables A_i avec un vecteur de coefficients d'effets β et un vecteur d'effets aléatoires individuels η_i qui suit une loi gaussienne de moyenne nulle et de variance une matrice définie positive Ω . Un modèle exponentiel pour les effets aléatoires peut être utilisé qui permet d'assurer la positivité des estimateurs des paramètres pharmacocinétiques.

Le vecteur θ des paramètres de population est constitué des effets fixes (μ), des effets des covariables (β), de la matrice de variance-covariance Ω et des paramètres du modèle d'erreur (a et/ou b).

1.3.2 Les méthodes d'estimation

La plupart des méthodes utilisées dans les MNLEM reposent sur l'estimation par le maximum de vraisemblance (ML) qui maximise la probabilité d'observer les données en sachant le modèle. Cependant, du fait de la non linéarité de f en ses paramètres, la vraisemblance n'a pas de solution analytique. Une solution pour résoudre ce problème est l'approximation de la vraisemblance par linéarisation. Sheiner a développé les premières méthodes, FO (*First Order*) qui linéarise le modèle autour de la moyenne des effets aléatoires (i.e. zéro) et FOCE (*First Order Conditional Estimation*) reposant sur le même principe mais linéarisant le modèle autour des estimations des effets aléatoires. L'algorithme FOCE est implémenté dans le logiciel NONMEM (SHEINER et BEAL, 1998), très largement utilisé dans l'industrie pharmaceutique, et dans la fonction `nlme` du logiciel R (R DEVELOPMENT CORE TEAM, 2008). Ces méthodes ont cependant tendance à produire des estimateurs biaisés et peu précis en présence de forte variabilité individuelle et/ou lorsque le nombre d'observations par sujet croît moins vite que le nombre de sujets (GE et al., 2004; VONESH, 1996).

Une autre alternative pour estimer la vraisemblance sans linéarisation du modèle est la quadrature adaptative de Gauss (PINHEIRO et BATES, 2000) qui utilise des approximations des intégrales contenues dans la vraisemblance. Cette méthode, mise en œuvre dans le logiciel SAS, est plus précise et moins biaisée mais nécessite des temps de calculs plus longs et peut se montrer moins stable.

Plus récemment, l'algorithme SAEM (*Stochastic Approximation Expectation Maximisation*) a été développé par KUHN et LAVIELLE (2005). Cette méthode est une version stochastique de l'algorithme EM (*Expectation Maximisation*) où les paramètres individuels sont considérés comme les valeurs manquantes. L'étape *expectation* de

l'algorithme SAEM se décompose en une première étape de simulation de paramètres individuels par chaînes de Markov de Monte Carlo en utilisant les valeurs courantes des paramètres puis en une étape d'estimation des statistiques suffisantes de la vraisemblance des données complètes. La maximisation de ces statistiques suffisantes permet d'obtenir de nouvelles valeurs des paramètres, utilisés à la première étape de l'itération suivante jusqu'à convergence de l'algorithme. L'algorithme SAEM ne nécessite donc pas de linéarisation. Cet algorithme est implémenté dans le logiciel MONOLIX (LAVIELLE, 2008), dans le logiciel NONMEM depuis la version 7, dans S-ADAPT depuis la version 1.56 et dans le logiciel R (package saemix) depuis juillet 2011. Dans cette thèse, nous avons utilisé l'algorithme SAEM implémenté dans MONOLIX (version 3.2 et puis 4.1).

Après estimation des paramètres de population, des estimations Bayésiennes empiriques des paramètres individuels du sujet i peuvent être obtenues à partir de la distribution *a priori* des paramètres et de la vraisemblance *a posteriori* pour ce sujet.

1.3.3 Le calcul de l'incertitude des paramètres

En modélisation, on s'intéresse aux paramètres de régression, et on voudrait la plupart du temps préciser la distribution de l'estimateur des paramètres. L'incertitude associée à l'estimation des paramètres est, en effet, un des critères principaux utilisés dans l'évaluation du modèle. Cette incertitude est souvent simplement résumée par l'erreur standard (SE).

Les SE sont obtenues après l'estimation des paramètres du MNLEM à partir de l'inverse de la matrice d'information de Fisher (MF), la borne inférieure de la matrice de variance covariance de tout estimateur non biaisé selon l'inégalité de Rao-Cramer. Ce calcul est fondé sur l'approximation asymptotique supposant la normalité asymptotique des estimateurs. Du fait de la non linéarité du modèle, MF n'est pas une solution analytique et une approximation est donc obtenue, soit par linéarisation au premier ordre du modèle autour de ses effets aléatoires soit par une approche stochastique reposant sur le principe de Louis (LOUIS, 1982) qui relie les fonctions du gradient et hessienne de la vraisemblance observée et celles de la vraisemblance des données complètes.

1.4 L'approche bootstrap pour l'estimation de l'incertitude

Les estimations de la précision des paramètres peuvent être biaisées lorsque l'approximation asymptotique est inappropriée, par exemple lorsque la taille de l'échantillon est petite. Parfois, elles ne peuvent pas être obtenues lorsque le modèle est complexe ou le protocole de prélèvement est épars. L'approche bootstrap représente une

technique alternative pour estimer la SE des paramètres et ainsi fournir l'intervalle de confiance.

1.4.1 Le principe du bootstrap

Le bootstrap a été introduit par EFRON (1979) pour les observations indépendantes et identiquement distribuées (iid). L'idée princeps du bootstrap consiste à rééchantillonner avec remise les données observées de façon répétée pour construire la distribution d'un estimateur ou d'une statistique de test (EFRON, 1979; EFRON et TIBSHIRANI, 1994). Cette idée est basée sur le principe de *plug-in*. Soit un échantillon X_1, \dots, X_n de n observations iid selon une loi F inconnue et $\theta(F)$ un paramètre que l'on souhaite estimer. Selon le principe de *plug-in*, $\theta(F)$ est estimé par $\theta(\hat{F})$, obtenu en remplaçant F par la fonction de répartition empirique \hat{F} calculé sur l'échantillon. Si $\theta(\hat{F})$ ne peut pas être calculé directement, on peut l'approcher par $\theta^*(\hat{F}^*)$ obtenu à partir de B échantillons bootstrap $((X_{1b}^*, \dots, X_{nb}^*)_{b=1, \dots, B})$.

En effet, en régénérant un grand nombre d'échantillons analogues aux échantillons de départ, le rééchantillonnage par bootstrap permet d'approcher la distribution réelle et inconnue de l'échantillon originel. L'approximation de la distribution bootstrap dans le cas de la moyenne normalisée (centrée) est plus précise que l'approximation normale (SHAO et TU, 1995). Cependant, ces propriétés du bootstrap ne sont pas bonnes dans le cas de distribution à queues lourdes (HALL, 1990).

1.4.2 Les méthodes du bootstrap en régression linéaire

Considérons le modèle de régression linéaire simple suivant,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.4)$$

où \mathbf{y} est le vecteur de n observations de la variable dépendante, \mathbf{X} est la matrice de dimension $n \times k$ de variables indépendantes, $\boldsymbol{\beta}$ est le vecteur de dimension k de paramètres inconnus et $\boldsymbol{\epsilon}$ représente le vecteur de dimension n des erreurs résiduelles supposées indépendantes et identiquement distribuées selon une distribution gaussienne de moyenne nulle et de variance σ^2 ($\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$).

Après l'ajustement, les résidus sont estimés comme suit :

$$\epsilon_i = y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}; i = 1, \dots, n \quad (1.5)$$

où $\hat{\boldsymbol{\beta}}$ est l'estimateur des moindres carrés ordinaires des paramètres ($\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{y}$).

L'approche bootstrap permet de préciser la distribution de $\hat{\beta}$. Soit B le nombre d'échantillons bootstrap à tirer des données originales, un algorithme général du bootstrap en régression est le suivant :

1. Générer un échantillon bootstrap par rééchantillonnage à partir des données originales et/ou du modèle estimé
2. Obtenir les estimations pour tous les paramètres du modèle pour l'échantillon bootstrap
3. Répéter les étapes 1 et 2 B fois pour obtenir la distribution bootstrap des estimations des paramètres et puis calculer la moyenne/la médiane, l'écart-type et l'intervalle de confiance à 95% de cette distribution

Plusieurs types de bootstrap ont été proposés pour rééchantillonner les données (étapes 1) dans le cas de la régression linéaire : le bootstrap par paires, le bootstrap des résidus, le bootstrap paramétrique et le bootstrap externe (SHAO et TU, 1995; DAVISON et HINKLEY, 1997; MACKINNON, 2006; WEHRENS et al., 2000). Le bootstrap par paires consiste à rééchantillonner avec remise parmi les individus ; c'est la forme la plus naïve et la plus intuitive, et l'échantillon bootstrap est alors une série de paires (\mathbf{X}, \mathbf{y}) provenant de l'échantillon originel. Le bootstrap des résidus consiste à rééchantillonner les résidus obtenus après un ajustement du modèle. Les résidus peuvent être centrés avant tirage, et des pondérations ont également été proposées pour corriger une certaine sous-estimation de la variance. Les résidus rééchantillonnés ϵ^* seront ensuite additionnés avec les prédictions du modèle pour construire les données bootstrap : $y_i^* = \mathbf{X}_i\hat{\beta} + \epsilon^*$. Le bootstrap paramétrique reprend le principe du bootstrap des résidus, mais au lieu de rééchantillonner directement les résidus observés, on simule des résidus dans la distribution estimée, par exemple la distribution normale. Le bootstrap externe consiste à tirer des résidus dans une distribution F dont les trois premiers moments sont connus. Différents choix de F ont été proposés. Lorsque la variance de l'erreur de mesure n'est pas constante, il est naturel de bootstrapper de préférence sur les résidus standardisés par la variance de chaque observation (par exemple, pour une variance proportionnelle à l'observation).

1.4.3 Les méthodes du bootstrap dans les modèles non linéaires à effets mixtes

Dans les MNLEM, trois niveaux de complexité supplémentaires apparaissent par rapport au bootstrap simple : la présence de données répétées chez un sujet, l'hétéroscédasticité et la non-linéarité du modèle. Les méthodes de bootstrap classiques développées en régression linéaire simple devraient être modifiées pour tenir compte des caractéristiques des MNLEM (DAS et KRISHEN, 1999). Elles doivent respecter le "vrai" processus générateur de données (*data generating process*, DGP) (HALIMI, 2005) en

généralisant deux niveaux de variabilité : la variabilité inter-sujet et la variabilité résiduelle. Le rééchantillonnage des effets aléatoires obtenus par l'approche bayésienne (des estimations bayésiennes empiriques) peut être couplé avec le rééchantillonnage des résidus (HALIMI, 2005; DAS et KRISHNAN, 1999; VAN DER LEEDEN et al., 1997; WU et ZHANG, 2002). Le bootstrap par paires peut être combiné avec le bootstrap des résidus (MACKINNON, 2006). Le rééchantillonnage peut être stratifié pour tenir compte de particularités du jeu de données, comme par exemple lorsque le jeu de données originel combine des patients ayant des données riches et des patients ayant des données éparpillées (ou limitées) ; une stratification est alors fréquemment utilisée pour éviter de n'inclure dans le jeu de données bootstrap que des patients en données éparpillées.

Dans le domaine de PK/PD de population, le bootstrap a été proposé comme une méthode de validation interne avancée (ETTE, 1997). La plupart des applications PK/PD utilisent la méthode du bootstrap pour évaluer le modèle final en comparant les estimations asymptotiques des paramètres obtenues pour les données originales avec celles obtenues par la méthode du bootstrap et estimer leurs erreurs standards et/ou les intervalles de confiance (PARKE et al., 1999; YAFUNE et ISHIGURO, 1999). Bien qu'assez fréquemment mis en œuvre, le bootstrap par paires n'a pas été formellement évalué, ni comparé de façon systématique avec d'autres alternatives en PK/PD tels que le bootstrap des résidus et le bootstrap paramétrique qui semblent le mieux approcher le vrai DGP.

1.5 Objectifs de la thèse

Du fait de la complexité des propriétés PK/PD des médicaments anti-angiogéniques, nous avons voulu développer des modèles mécanistiques pour étudier le mécanisme d'action de ces nouveaux médicaments. Nous nous sommes également intéressées à l'étape d'évaluation du modèle, particulièrement l'incertitude d'estimation des paramètres obtenus par l'approche bootstrap. Les objectifs principaux de cette thèse sont d'une part de modéliser la PK et la PD de l'affibercept (Zaltrap®), un nouvel anti-angiogénique et d'autre part de développer et comparer différentes approches bootstrap dans le cadre du modèle non linéaire à effets mixtes par l'intermédiaire d'études de simulation.

Les résultats de la construction du modèle PK pour l'affibercept libre et lié chez les volontaires sains sont présentés dans la partie 2.1. Nous avons appliqué ce modèle aux données issues de 9 essais cliniques chez les patients atteints de cancer, étudié l'influence de facteurs physiopathologiques sur la PK de l'affibercept et évalué différents schémas d'administration par simulation. Ce travail est présenté dans la partie 2.2. Ce modèle PK nous a amené à développer un modèle PK/PD caractérisant l'inhibition de la croissance tumorale sous l'effet du complexe affibercept-VEGF (l'affibercept lié) chez des patients atteints du cancer colorectal métastatique dans la partie 2.3.

En parallèle, nous avons étudié par simulation les performances des différentes méthodes de bootstrap prenant en compte les caractéristiques des MNLEM avec deux niveaux de variabilité (inter-sujet et résiduelle) présentées au chapitre 3. La première simulation a été mise en place d’abord dans le modèle linéaire à effets mixtes, pour lequel des résultats théoriques existent, avec l’exemple de progression naturelle de la maladie de Parkinson. Ce travail est présenté dans la partie 3.1 et nous a permis d’éliminer les méthodes ayant de mauvaises performances et de sélectionner les trois meilleures méthodes pour les évaluer dans les MNLEM avec l’hétéroscédasticité. Cette deuxième simulation a été basée sur les données des concentrations de l’affibercept libre issues de deux essais cliniques. Ce travail est présenté dans la partie 3.2. Enfin, nous avons appliqué la méthode bootstrap par paires sur les modèles PK d’affibercept libre et lié développés précédemment.

Chapitre 2

Développement des modèles mécanistiques pour la PK et la PD d'aflibercept dans le traitement du cancer

2.1 Développement d'un modèle mécanistique pour la pharmacocinétique d'aflibercept libre et lié chez les volontaires sains

2.1.1 Résumé

L'angiogenèse, la croissance de nouveaux vaisseaux sanguins à partir de vaisseaux préexistants, joue un rôle crucial dans la croissance des tumeurs malignes. Elle est médiée notamment par le facteur de croissance vasculaire endothélial (VEGF), cible thérapeutique de nouveaux médicaments anti-angiogéniques comme l'aflibercept (Zaltrap[®], développé conjointement par Regeneron et Sanofi). Il s'agit d'une protéine de fusion comportant des domaines des récepteurs VEGFR-1 et VEGFR-2 et une portion Fc de l'immunoglobuline G1 humaine. Il bloque le VEGF-A et le VEGF-B ainsi que le facteur de croissance placentaire (PlGF) et donc l'angiogenèse.

Dans cette première étude, nous avons développé un modèle PK mécanistique pour l'aflibercept afin de caractériser sa liaison au VEGF et ses propriétés PK chez les volontaires sains. Les données utilisées sont issues de deux études cliniques de phase I avec l'aflibercept administré en une seule perfusion intraveineuse de 1, 2 ou 4 mg/kg. Les concentrations plasmatiques de l'aflibercept libre et lié au cours du temps ont été analysées par l'approche de modélisation non linéaire à effets mixtes. Dans un premier temps, nous avons modélisé les données de l'aflibercept libre tout seul et puis dans un second temps, nous les avons modélisées conjointement avec les données de l'aflibercept lié. Les modèles TMDD et ses approximations ainsi que le modèle de liaison linéaire ont été utilisés pour décrire l'évolution des deux entités. Nous avons estimé les paramètres PK ainsi que leur variabilité inter-sujet en utilisant l'algorithme SAEM implémenté dans

le logiciel MONOLIX 3.1. Nous avons également pris en compte la présence des données sous la limite de quantification pour l'affibercept lié (constituant 32.5% des données) en utilisant l'algorithme SAEM étendu (SAMSON et al., 2006). Le test du rapport de vraisemblance a été utilisé pour distinguer les modèles emboîtés, basé sur la différence en log-vraisemblance (-2 LL). Pour les modèles non emboîtés, la sélection des modèles a été basée sur le critère d'information bayésien (BIC). Le meilleur modèle est celui qui a la plus petite valeur de BIC.

Le meilleur modèle trouvé pour l'affibercept libre et lié est une approximation de type Michaelis-Meten (MM) du modèle TMDD comportant deux compartiments pour l'affibercept libre, un compartiment pour l'affibercept lié et une liaison irréversible de l'affibercept libre au VEGF dans le compartiment périphérique. Ce modèle décrit correctement la PK de l'affibercept. Il caractérise bien son mécanisme d'action sous-jacent et sa liaison non linéaire au VEGF. A notre connaissance, c'est le premier modèle PK de population publié basé sur le mécanisme d'action pour un médicament anti-VEGF.

Cette analyse a fait l'objet d'un article publié dans la revue *British Journal of Clinical Pharmacology*.

2.1.2 Article 1 (publié)

A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects

Hoai-Thu Thai,^{1,2} Christine Veyrat-Follet,¹ Nicole Vivier,¹
Catherine Dubruc,¹ Gerard Sanderink,¹ France Mentré² &
Emmanuelle Comets²

¹Global Metabolism and Pharmacokinetics Department, Sanofi-aventis and ²UMR738, INSERM, University Paris Diderot, Paris, France

Correspondence

Ms Hoai-Thu Thai, UMR738, INSERM, University Paris Diderot, 16 rue Henri-Huchard, 75018 Paris, France.
Tel.: +33 1 5727 7538
Fax: +33 1 5727 7521
E-mail: hoai-thu.thai@inserm.fr

Keywords

aflibercept, MONOLIX, population pharmacokinetics, target-mediated drug disposition, VEGF

Received

1 December 2010

Accepted

7 May 2011

Accepted Article

17 May 2011

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT

- Anti-angiogenic drugs have been developed as an effective therapeutic strategy for inhibiting tumour growth. However, their pharmacokinetics (PK) and their ligand inhibition properties have not been well characterized. The binding to a circulating target, such as vascular endothelial growth factor (VEGF), makes the PK of these drugs more complex.

WHAT THIS STUDY ADDS

- The underlying mechanism of disposition of aflibercept, where a saturable and high affinity binding of aflibercept to VEGF was adequately characterized by the Michaelis–Menten approximation of a target-mediated drug distribution model. To our knowledge, this is the first published mechanism-based population PK model for an anti-VEGF drug.

AIM

Aflibercept (VEGF-Trap), a novel anti-angiogenic agent that binds to VEGF, has been investigated for the treatment of cancer. The aim of this study was to develop a mechanism-based pharmacokinetic (PK) model for aflibercept to characterize its binding to VEGF and its PK properties in healthy subjects.

METHODS

Data from two phase I clinical studies with aflibercept administered as a single intravenous infusion were included in the analysis. Free and bound aflibercept concentration–time data were analysed using a nonlinear mixed-effects modelling approach with MONOLIX 3.1.

RESULTS

The best structural model involved two compartments for free aflibercept and one for bound aflibercept, with a Michaelis–Menten type binding of free aflibercept to VEGF from the peripheral compartment. The typical estimated clearances for free and bound aflibercept were 0.88 l day⁻¹ and 0.14 l day⁻¹, respectively. The central volume of distribution of free aflibercept was 4.94 l. The maximum binding capacity was 0.99 mg day⁻¹ and the concentration of aflibercept corresponding to half of maximum binding capacity was 2.91 µg ml⁻¹. Interindividual variability of model parameters was moderate, ranging from 13.6% (V_{max}) to 49.8% (Q).

CONCLUSION

The present PK model for aflibercept adequately characterizes the underlying mechanism of disposition of aflibercept and its nonlinear binding to VEGF.

Introduction

Angiogenesis, the development of new blood vessels from pre-existing vasculature, participates in a variety of physiological processes and disease states [1]. Its critical role in tumour development and progression was established 30 years ago [2]. This discovery brought a new effective approach called anti-angiogenic therapy to cancer treatment, which consists in limiting blood supply to tumours by preventing angiogenesis. The development of new agents has attracted many researchers' interest in the pharmaceutical industry. To date, the best characterized and most highly validated anti-angiogenic approach involves targeting the vascular endothelial growth factor (VEGF) pathway [3].

Vascular endothelial growth factor is the most potent pro-angiogenic growth factor, promoting the formation of blood vessels which are required for both normal and neoplastic tissue growth [1, 4]. VEGF binds to two high-affinity receptors (VEGFR-1 and VEGFR-2) on endothelial cells. This binding activates the intrinsic tyrosine kinase activity of their cytodomains, initiating intracellular signalling. VEGF is expressed in a large variety of malignant tumours, such as tumours of the breast, brain, lung and gastrointestinal tract [5]. Blockade of the VEGF pathway is therefore an effective therapeutic strategy for inhibiting tumour growth [4–6].

Aflibercept (also called VEGF-Trap, Regeneron Pharmaceuticals/Sanofi-aventis research) is a novel anti-angiogenic agent that binds to VEGF with a 1:1 ratio and prevents it from interacting with its receptors. A recombinant fusion protein consisting of the second Ig domain of VEGFR-1 and the third Ig domain of VEGFR-2 fused to the Fc portion of human immunoglobulin IgG1, it has a higher affinity for VEGF-A (K_d *in vitro* = 0.5 pM) than current anti-VEGF monoclonal antibodies [7–9]. Aflibercept also binds to VEGF-B and Placental Growth Factor (PlGF), which may be advantageous in some settings, such as malignant ascites where PlGF may mediate vascular permeability [9].

Based on the mechanism of action, this drug undergoes a target-mediated drug disposition (TMDD), a term used to describe the phenomenon in which drug is bound with high affinity to its pharmacologic target such that this interaction is reflected in the pharmacokinetic properties of the drug. A general PK model for drugs exhibiting TMDD has been developed by Mager *et al.* [10, 11]. This model describes the elimination pathway of drug plasma concentrations as the combination of first-order elimination from the central compartment and specific target binding clearance followed by internalization of the drug–target complex. It also characterizes the turnover of the target. The full TMDD model is complex and generally overparameterized. The more information we have about free drug, bound drug and the target, the more TMDD model components and parameters can be adequately identified, although it is yet unclear which elements should be measured to estimate all the parameters in the full TMDD

model. In order to overcome this problem, several simpler forms of the TMDD model were proposed [12, 13]. There are mainly three approximations: quasi equilibrium (QE), quasi steady-state (QSS) and Michaelis–Menten (MM). The QE approximation is based on the assumption that the drug–target binding is much faster than all other system processes. If the rate of elimination of the complex is not negligible, the QE approximation is replaced by the QSS approximation assuming that the drug–target complex concentration changes more slowly than the binding and internalization process. The MM approximation describes the system when the target concentration is small relative to the free drug concentration and the dosing regimens result in the target being fully saturated [13].

Pharmacokinetics (PK) of aflibercept were investigated in healthy subjects after single intravenous (i.v.) doses of 1 to 4 mg kg⁻¹ and a single subcutaneous (s.c.) dose of 2 mg kg⁻¹, in two phase 1 clinical studies as part of the drug's clinical development. Both free and bound aflibercept concentrations were assayed. The objective of this analysis was to develop a mechanism-based PK model for aflibercept in order to characterize its binding to VEGF and its pharmacokinetic properties in healthy subjects. The influence of covariate was not assessed in this study due to the limited number of individuals and their healthy status.

Methods

Study design

The data for the population PK analysis were collected from two phase 1, monocentric and randomized studies which were both carried out in healthy male subjects to assess the PK of aflibercept. The studies were approved by the independent ethics committees (Pharma-Ethics, South Africa and Ethik-Kommission der Landesärztekammer Baden-Württemberg, Germany). They were performed according to recommendations of the 18th World Health Congress (Helsinki, 1964) and all applicable amendments. The volunteers gave their written informed consent after full explanation of all procedures involved in the studies.

Study 1 was a placebo-controlled, single-dose, sequential ascending-dose study. Forty-eight subjects were enrolled in this study and equally divided into four groups: one group receiving placebo and three groups receiving a single dose of 1, 2 or 4 mg kg⁻¹ of aflibercept, respectively, administered as a 1-h i.v. infusion.

Study 2 was an open-label, single-dose, crossover study. Two groups of 20 subjects were included in the study. The first group received a single i.v. dose of 2 mg kg⁻¹ in the first period, with a 2-month follow-up period, followed by a single s.c. dose of 2 mg kg⁻¹ in the second period. The second group received the s.c. dose first, then the i.v. dose.

The data from s.c. administration in the crossover study (study 2) were removed from the analysis because this route of administration was not pursued in the subsequent

clinical development. Moreover, in this study, a carry-over effect was found and should have been taken into account in the modelling, but this required the modelling of the s.c. route. To avoid it and to work on homogenous data, only the i.v. infusion data in the first period were used in the population analysis.

Blood sampling schedules

In study 1, blood samples (4 ml) were taken at the following times: pre dose, 1 (end of infusion), 2, 4, 6, 8, 12 and 24 h post start of administration on day 1, then on days 8, 15, 22, 29, 36 and 43 at the same morning time corresponding to 2 h after the start of infusion on day 1.

In study 2, blood samples (4 ml) were taken at the following times: pre dose, 1, 2, 4, 6 and 8 h post start of administration on day 1, then on days 2, 3, 5, 8, 15, 2 and 43 of each period at the same morning time corresponding to 2 h after the start of infusion on day 1.

Assay method

For both studies, free and bound aflibercept plasma concentrations were measured in all samples collected at Regeneron Pharmaceuticals, Inc. using enzyme-linked immunosorbent assay (Elisa) method. Blood samples were collected in tubes (containing 1 ml of citrate buffer, sodium citrate, and 4.2 mg of citric acid) and were centrifuged at 2000 g for 15 min at room temperature. Plasma was stored at -20°C until analysed.

In the assay of free aflibercept, human VEGF165 was initially adsorbed to the surface of a polystyrene solid support to capture the free aflibercept in the samples. A mouse monoclonal antibody (reporter antibody), specific to an epitope on the VEGFR1 domain of aflibercept, was then bound to the immobilized complex and an enzyme-linked antibody (peroxidase-conjugated Affinipure goat anti-mouse IgG Fc- γ) was bound to the immobilized mouse monoclonal complex. A luminal-based substrate specific for peroxidase was added to achieve a signal intensity that was directly proportional to the concentration of free aflibercept. The limit of quantification of free aflibercept was 15.6 ng ml^{-1} . The calibration curves ranged from 100 ng ml^{-1} to 1.56 ng ml^{-1} in twofold serial dilution. The limit of quantification of free aflibercept was 15.6 ng ml^{-1} . The inter-day accuracy and precision ranged from 92.24% to 103.09% and 1.05% to 16.18%, respectively. The intra-day accuracy and precision ranged from 106.36% to 109.90% and 9.56% to 13.68%, respectively.

In the assay of bound aflibercept, human VEGF165 was replaced by a non-blocking goat anti-human VEGF antibody for capturing the bound aflibercept in subject samples. The rest of the procedure was similar to the one used in the assay of free aflibercept. The limit of quantification for bound aflibercept was 43.9 ng ml^{-1} . The calibration curves ranged from 100 ng ml^{-1} to 8.78 ng ml^{-1} in a 1.5 serial dilution. The inter-day accuracy and precision ranged from 93.85% to 110.19% and 0.38% to 16.17%, respec-

tively. The intra-day accuracy and precision ranged from 115.70% to 123.34% and 1.09% to 1.58%, respectively.

Population PK analysis

The population PK analysis was performed using the MONOLIX program (version 3.1) implementing the SAEM algorithm. The model control files were written using MLXTRAN script. The early concentrations of bound aflibercept were often below the limit of quantification, thus the censored data of bound aflibercept, representing 32.5% of data, were taken into account and used for model development using the extended SAEM algorithm implemented in MONOLIX as an exact maximum likelihood estimation method [14]. In this algorithm, the left-censored data are simulated in a right-truncated Gaussian distribution, instead of being imputed by the LOQ value or half of the LOQ value. The data of bound aflibercept contains also some observed concentrations which were reported with values below the LOQ reported for most of the data (see bottom left plot in Figure 4).

The database included a total of 56 subjects, with 36 subjects receiving treatments from study 1 and 20 subjects receiving i.v. infusions in the first period from study 2. With respect to the law of mass action, the concentrations of bound aflibercept were converted into equivalent concentrations of free aflibercept by multiplying them with 0.717, the ratio of molecular weights between free and bound aflibercept. The units of free aflibercept and bound aflibercept concentrations were $\mu\text{g ml}^{-1}$ and $\mu\text{g eq ml}^{-1}$ respectively.

Pharmacokinetic structural model

The following strategy was used to develop the model. Free aflibercept concentration–time data were first modelled alone. Then, bound concentration–time data were included for simultaneous modelling.

The structural model for free aflibercept was developed by testing the following models: two-compartment or three-compartment models with first-order and/or MM elimination.

In the next step, we developed a structural model including bound aflibercept. The TMDD model with association and dissociation rate constants (k_{on} and k_{off}), reduced approximate TMDD models and other simpler models with linear binding constants were used to describe the joint evolution of the two entities.

Statistical model

Denoting f the function describing the model, the statistical model for observed concentration C_{ij} of subjects i for sampling time t_{ij} is:

$$C_{ij} = f(\theta_i, t_{ij}) + \varepsilon_{ij}$$

where θ_i is the vector of parameters of subject i and ε_{ij} is the residual error.

The errors ε_{ij} were assumed to be independent and normally distributed with a null mean and a heteroscedastic variance σ^2_{ij} , which was modelled using a combined additive and proportional model:

$$\sigma^2_{ij} = (\sigma_a + \sigma_p f(\theta_i, t_{ij}))^2$$

where σ_a and σ_p are additive and proportional coefficients of the residual error model respectively.

Two alternative residual error models, proportional model ($\sigma_a = 0$) and additive model ($\sigma_p = 0$) were also evaluated for the residual variability.

The interindividual variability on all parameters was modelled with an exponential model, e.g. for CL:

$$CL_i = CL \cdot \exp(\eta_{i,CL})$$

where $\eta_{i,CL}$ denotes the random effect in subject i , CL_i the individual clearance parameter and CL the typical value of the population. The use of an exponential model implies a log-normal distribution for the parameters. The η s (e.g. $\eta_{i,CL}$) are zero mean random variables with variance ω^2 (e.g. for CL , $\omega^2 CL$). The ω^2 s represent the variance of the random effects. The elements of the interindividual variance-covariance matrix, Ω , was modelled as diagonal, e.g. assuming no covariance between the η s.

The following strategy was used for model development. First, the structural PK models were developed with a combined residual error model and interindividual variability on all parameters. Then, the residual error models were evaluated for the selected structural model. Finally, the interindividual variability on each parameter was tested for significance, and the non-significant variability components were removed one by one starting with the smallest and least significant estimate.

The log likelihood (LL) was computed using importance sampling. The likelihood ratio test was used to discriminate between nested models through the difference in log likelihood ($-2 LL$). A P -value of 0.05 was considered statistically significant. For non-nested models, the model selection was based on the Bayesian information criterion (BIC). The better model is the one with a smaller value of BIC [15].

Model evaluation

Internal evaluation of the model was based on goodness-of-fit (GOF) plots, including plots of observations vs. individual and population predictions and plots of normalized prediction distribution error (NPDE) [16]. A visual predictive check (VPC) was used to assess model predictive performance, based on the simulation of 500 data sets.

Results

Data

The database for population PK analysis contained data from 56 healthy subjects involved in the two phase 1 clinical

studies. A total of 1476 concentrations were used for model building: 732 concentrations of free aflibercept and 744 concentrations of bound aflibercept of which 242 (32.5%) were below the quantification limit (LOQ = 43.9 ng ml⁻¹ or 0.0314 µg eq ml⁻¹).

The pooled concentrations of free aflibercept and bound aflibercept plotted vs. time are presented in Figure 1.

The time course of free aflibercept after i.v. infusion of doses of 1, 2 and 4 mg kg⁻¹ in semi-log scale suggests a bi-exponential decline with a rapid phase of distribution followed by a prolonged terminal phase of elimination regardless of the dose. The time course of bound aflibercept suggests a saturable binding phenomenon with the same observed plateau for the two higher doses. The peaks of complex occurred sooner for 1 mg kg⁻¹ and later for doses of 2 and 4 mg kg⁻¹ (around 21 days for both doses).

Free aflibercept modelling

Several models were proposed to characterize the kinetics of free aflibercept. In the two-compartment model with first-order elimination, the distribution of *post hoc* clearance for three doses showed a dose-dependent clearance, confirming the nonlinear disposition already observed during non-compartmental analysis. The two-compartment model with MM elimination was then developed and showed a better description of data with a decrease of 33.61 points in $-2 LL$ value. The three-compartment models with first-order or MM elimination were also evaluated. They provided a slight decrease in $-2 LL$ compared with the two-compartment models but the fitting did not show a significant improvement. The two-compartment model with MM elimination was therefore selected as an adequate structural model for free aflibercept.

The best residual error for this model was the combined additive and proportional error. The interindividual variability on K_m was not significant.

Free and bound aflibercept modelling

The binding of aflibercept to VEGF and the nonlinear kinetics of free aflibercept suggests that this drug has target-mediated drug disposition properties. Therefore, the TMDD model with association and dissociation rate constants (k_{on} and k_{off}) reduced approximate TMDD models and other simpler models with linear binding constants were tested. The MM approximation of the TMDD model developed by Gibiansky and colleagues [13] was found to be the best approach to describe the kinetics of both free and bound aflibercept, while others did not fit the concentrations of bound correctly and/or had overparameterization issues. In all of these models, the dissociation rate k_{off} was very small compared with other constants and very badly estimated.

The nonlinear central and peripheral binding of this MM approximate TMDD model were tested and compared.

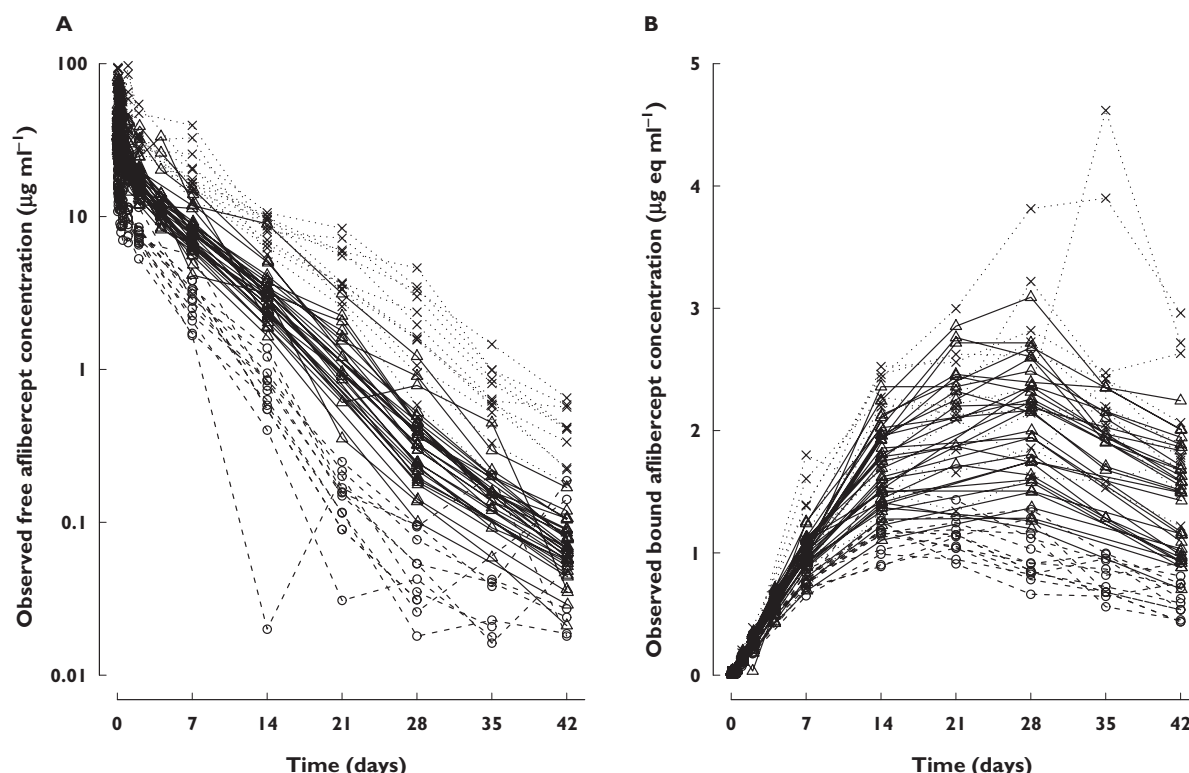


Figure 1

Observed concentrations of (A) free aflibercept ($\mu\text{g ml}^{-1}$) and (B) bound aflibercept ($\mu\text{g eq ml}^{-1}$) vs. time

The nonlinear peripheral binding model was found to describe the data better than the central binding one with a smaller value of BIC (2160 vs. 2703) and therefore was retained as the structural model for free and bound aflibercept.

The final structural model relating free and bound aflibercept is shown in Figure 2.

Free aflibercept in plasma distributes first to tissues then binds to VEGF to form the complex. Binding to VEGF follows the law of mass action and can be characterized by a nonlinear form with MM constants (V_{\max} and K_m). The bound aflibercept (complex) is assumed to be directly eliminated through internalization (k_{int}) and not through the dissociation rate constant (k_{off}) which gives back free aflibercept and free VEGF. The concentration of free aflibercept in central compartment (C_p), in the tissue compartment (C_t) and the concentration of bound aflibercept (C_b) are described by means of differential equations:

$$\frac{dC_p}{dt} = \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \quad (1)$$

$$\frac{dC_t}{dt} = k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{1}{V_t} \cdot \frac{V_{\max} \cdot C_t}{K_m + C_t} \quad (2)$$

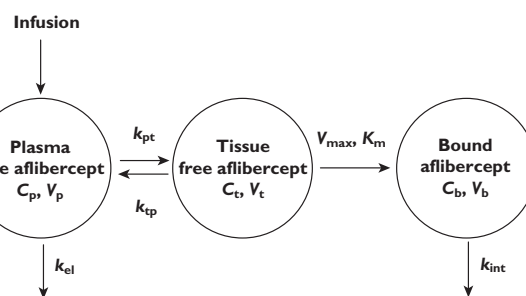


Figure 2

Proposed structural models for free and bound aflibercept (two compartments for free aflibercept, one compartment for bound aflibercept) with binding to VEGF occurring in the peripheral compartment

$$\frac{dC_b}{dt} = \frac{1}{V_b} \cdot \frac{V_{\max} \cdot C_t}{K_m + C_t} - k_{\text{int}} \cdot C_b \quad (3)$$

in which k_{el} is the first-order elimination rate constant of free aflibercept (day^{-1}) from the central compartment, k_{tp} and k_{pt} are the first-order rate constants between central

and peripheral compartment (day^{-1}), k_{int} is the first-order rate constant of bound aflibercept internalization (day^{-1}), V_{max} is the maximum binding capacity (mg day^{-1}), K_m is the concentration of free aflibercept corresponding to half of maximum binding capacity ($\mu\text{g ml}^{-1}$), V_p is the central volume of distribution of free aflibercept (l), V_t is the peripheral volume of distribution of free aflibercept and V_b is the volume of distribution of bound aflibercept (l). More details concerning the derivation of the model equations and its relationship to other TMDD models proposed in the literature can be found in the Appendix.

In this model, the volume of bound aflibercept (V_b) and the maximum binding capacity (V_{max}) were however highly correlated with a correlation coefficient higher than 0.95, suggesting identifiability issues. In order to prevent this problem, the value of V_b was assumed to be equal to the value of central volume of distribution of free aflibercept (V_p), corresponding to the hypothesis that free and bound aflibercept are distributed within the same space in the plasma compartment. To do so, V_b was fixed to the population value of V_p with no interindividual variability, instead of being estimated.

The best residual error model of bound aflibercept was the proportional model, while that of free aflibercept was the combined additive and proportional error model found in the first step. A combined error model was also attempted. However, the contribution of the additive error was negligible and was therefore discarded. The interindividual variability on the internalization rate constant (k_{int}) was found to be small (11.9%) and badly estimated. The likelihood ratio test demonstrated that removing this variability from the model did not significantly change the fit.

The estimated parameter values for the best model of free and bound aflibercept are presented in Table 1.

All model parameters were well estimated with relative standard errors (RSE) $<20\%$. The population estimate of clearance for free aflibercept was 0.88 l day^{-1} and the intercompartmental rate for bound aflibercept was 0.028 day^{-1} , resulting in a clearance of 0.14 l day^{-1} . V_p for free aflibercept was 4.94 l . The maximum binding capacity was

0.99 mg day^{-1} and the estimated concentration of free aflibercept corresponding to half of maximum binding capacity was low ($2.91 \mu\text{g ml}^{-1}$).

Interindividual variability for random effects associated with model parameters was moderate, ranging from 13.6% (V_{max}) to 49.8% (Q). Residual errors were low for both free and bound aflibercept with proportional parts of 17.1% and 12.6% respectively. The additive error of free aflibercept was also small compared with the observed concentrations.

The goodness-of-fit plots of the best joint model are shown in Figure 3. The plots of observations vs. population predictions and observations vs. individual predictions indicated that the model adequately described the observations over the dose range. The normalized prediction distribution error plots, NPDE vs. time and NPDE vs. predictions, showed a symmetric distribution around zero, except for the early times and the small concentrations of bound aflibercept. This bias was caused by omitting the NPDE corresponding to below the quantification limit (BQL) observations. For BQL data, the metric NPDE has not yet been developed.

The total number of BQL concentrations of bound aflibercept predicted by the model was 251 (33.73%), 231 of which corresponded to observed BQL data. This shows the good agreement between model prediction and observation for BQL data.

Examples of individual fits taken from two studies with three doses are shown in Figure 4. The model described the observed concentrations of free and bound aflibercept for all subjects quite well.

Figure 5 presents the visual predictive checks for both free and bound aflibercept. The 80% prediction interval and median were obtained by simulating 500 data sets under the final model. In addition, we obtained the 90% prediction intervals around the median and the upper and lower boundaries of the interval. Concentrations lower than the LOQ simulated for free aflibercept at the end of treatment were replaced by the LOQ ($0.0156 \mu\text{g ml}^{-1}$) in order to obtain a lower limit for the VPC plots on a

Table 1

1-Parameter estimates for the best free and bound model

Fixed effects Parameter	Estimate (RSE%)	Interindividual variability ω (%) (RSE%)	Residual variability	σ_a ($\mu\text{g ml}^{-1}$) (RSE%)	σ_p (%) (RSE%)
CL (l day^{-1})	0.88 (4.0)	28.0 (10)	Free aflibercept	0.05 (9.0)	17.1 (3.0)
V_p (l)	4.94 (4.0)	27.3 (10)	Bound aflibercept	–	12.6 (4.0)
Q (l day^{-1})	1.39 (9.0)	49.8 (14)			
V_t (l)	2.33 (7.0)	39.8 (14)			
V_b (l)	4.94 ($=V_p$)	–			
V_{max} (mg day^{-1})	0.99 (5.0)	13.6 (17)			
K_m ($\mu\text{g ml}^{-1}$)	2.91 (11)	45.6 (15)			
k_{int} (day^{-1})	0.028 (5.0)	–			

CL: clearance of free aflibercept from central compartment ($\text{CL} = k_{\text{el}} \cdot V_p$). Q: intercompartment clearance of free aflibercept ($Q = k_{12} \cdot V_t = k_{21} \cdot V_p$).

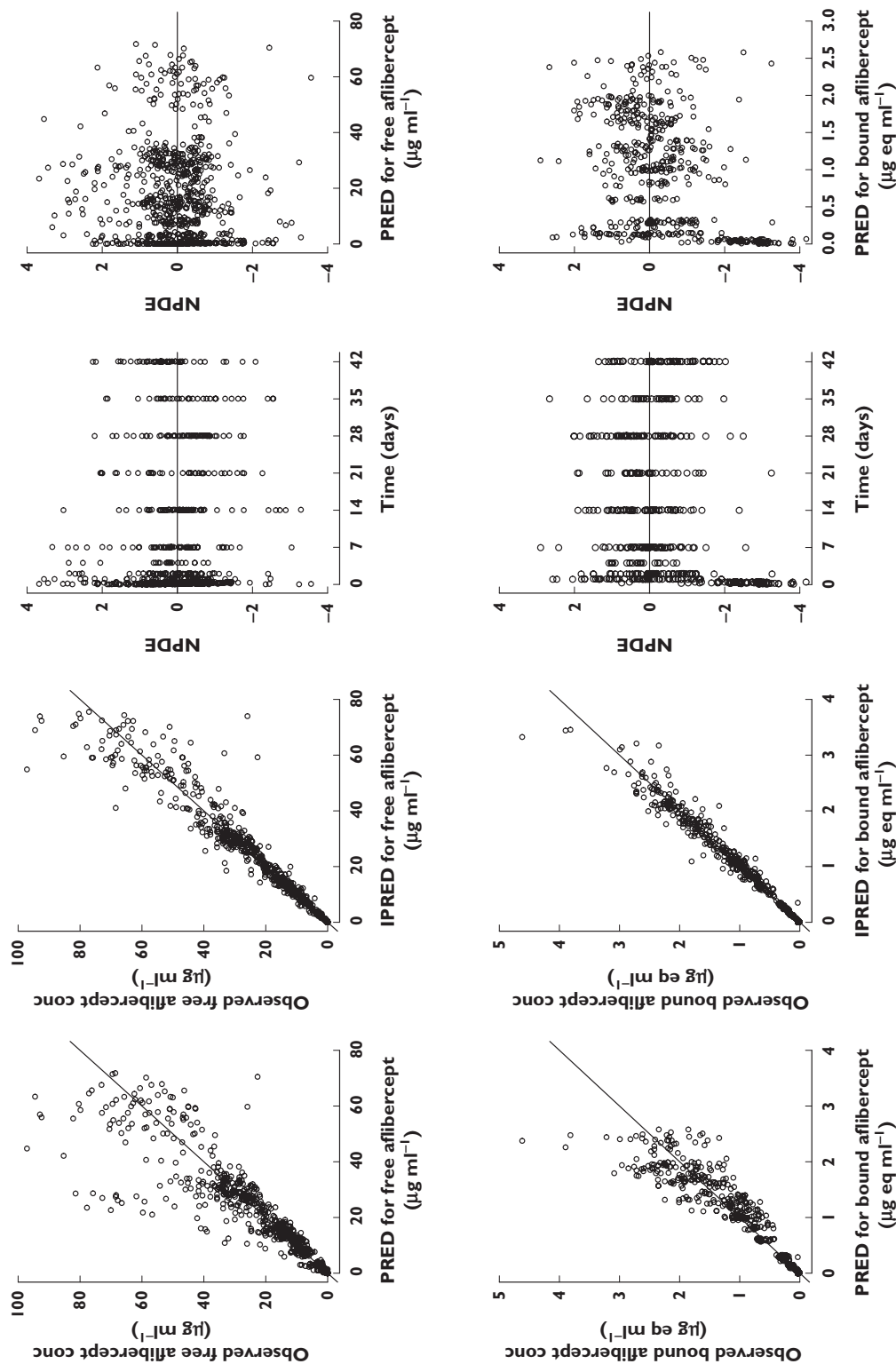


Figure 3

Diagnostic goodness-of-fit plots for the model of free and bound affibercept, showing observed vs. population predicted concentrations (PRED), observed vs. individual predicted concentrations (IPRED) and normalized prediction distribution error (NPDE) vs. population predicted concentrations. The lower limit of quantification for bound affibercept was 43.9 ng ml⁻¹ (0.0314 µg eq ml⁻¹). BQL observations were removed from the plots

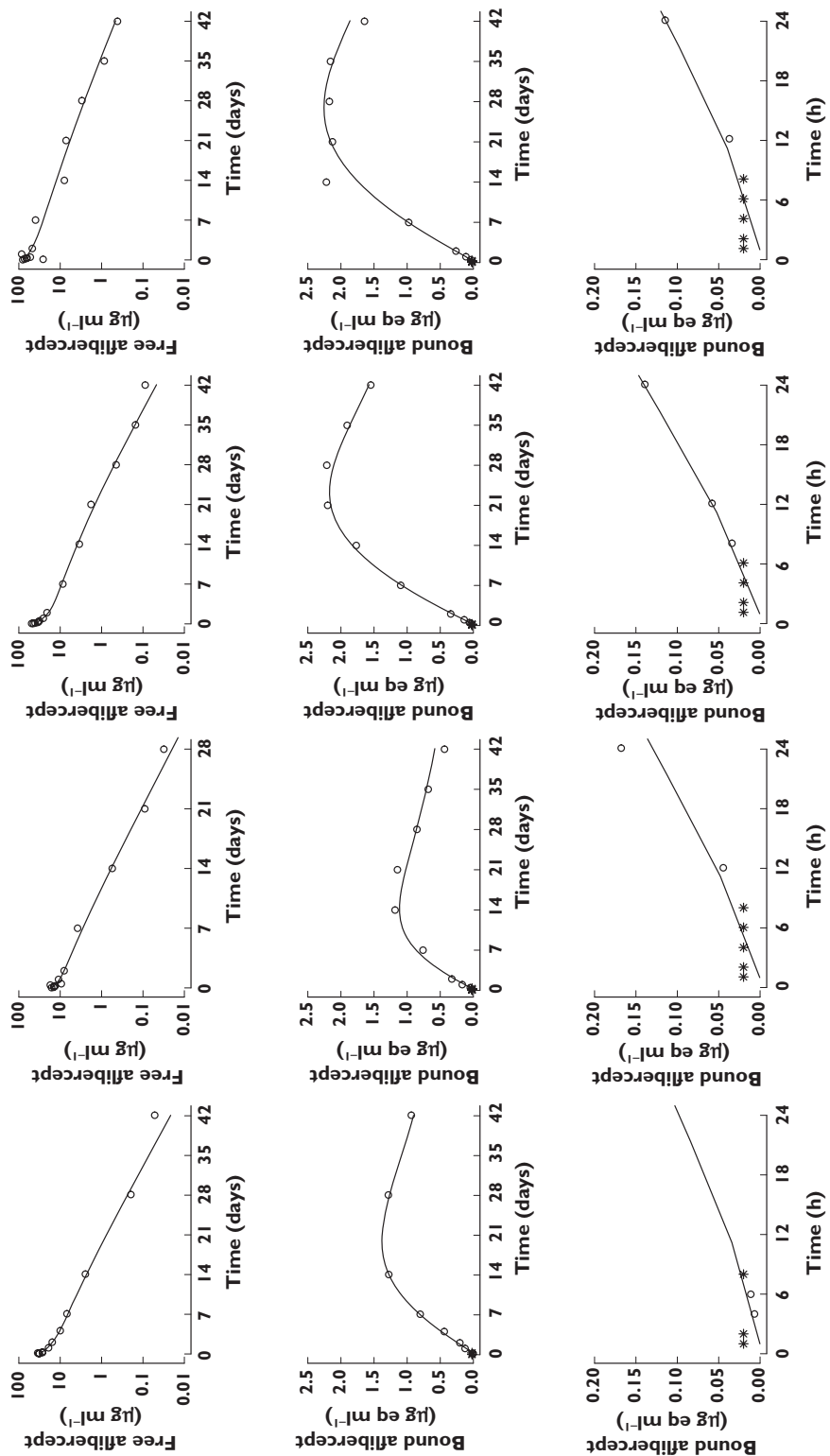


Figure 4

Examples of individual fits for four representative individuals. From left to right: study 1 with dose of 2 mg kg⁻¹, study 2 with dose of 1, 2, 4 mg kg⁻¹. Fits for free aflibercept are presented in the top, bound aflibercept in the middle, and the first 24 h fits of bound aflibercept at the bottom. Observed data are plotted using a circle (○) and BQL data are plotted using a star (*). The line (—) represents the prediction of model

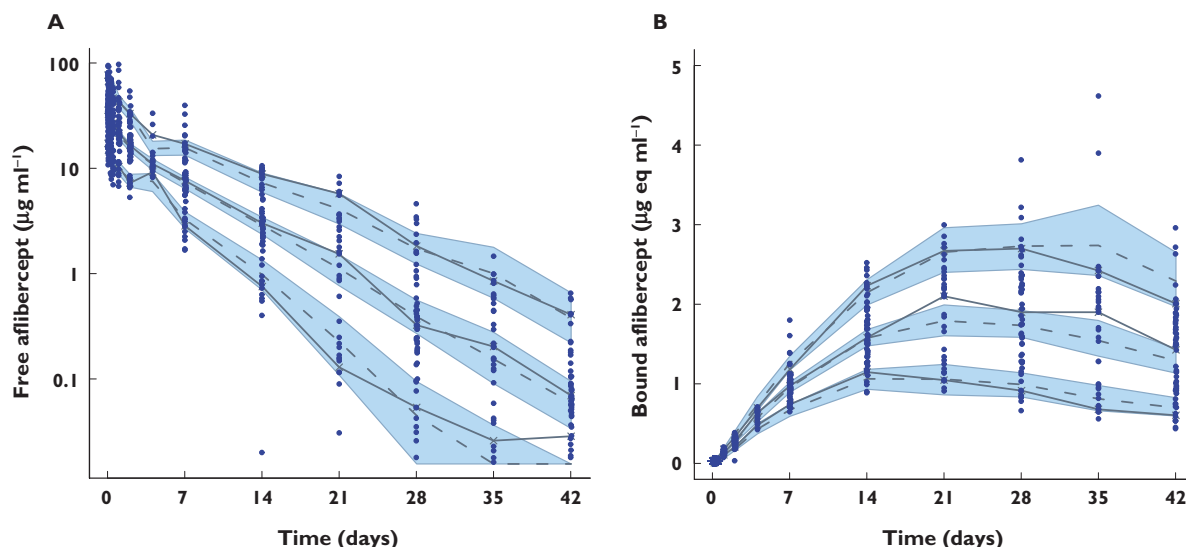


Figure 5

Visual predictive check (VPC) of free aflibercept in semi-logarithmic scale (A), and bound aflibercept in linear scale (B) for the final model. Observed data are plotted using a solid circle (●) and censored data are plotted using a star (*). The shaded area and the dotted lines represent the 90% prediction interval and the predicted median of 10th, 50th and 90th percentiles of simulated data ($n = 500$). The solid lines represent the 10th, 50th and 90th percentiles of observed data

semi-logarithmic scale. The slight misfit for free aflibercept seen at the last time point may be a consequence of the number of concentrations near the LOQ, since the median and upper boundary appear to be very well predicted by the model. For bound aflibercept, the model slightly underpredicted median concentration from day 21 onwards but correctly predicted the variability. Thus, the model described reasonably well the observed concentrations of both free and bound aflibercept.

Discussion

In this study, we report the development of a mechanism-based model to characterize the population PK of aflibercept after a single i.v. infusion of a 2 to 4 mg kg^{-1} dose to 56 healthy subjects. Free and bound aflibercept plasma concentrations were used in the modelling.

We first modelled free aflibercept concentrations alone. Using standard compartmental PK models, we observed a decrease of clearance for higher concentrations. This was consistent with previous results from noncompartmental analyses, which demonstrated that the volume of distribution of free aflibercept was low and its clearance was dose-dependent. When dealing with saturable kinetics, the most common model to consider is MM elimination kinetics, which was retained as the final model for free aflibercept during model development. Although such a model has been successfully applied to describe nonlinear PK of free aflibercept, it does not represent well the underlying

molecular events such as target binding, internalization and degradation of the drug. However, the modelling of free aflibercept helped us get general information about the model structure: two-compartment kinetics and dose-dependent clearance. More complex models combining linear and nonlinear elimination were also tested. They adequately fitted the data but were not retained as the final model of free aflibercept according to the statistical criteria.

The next step of modelling was then to add bound aflibercept data, which served as a marker of efficacy by representing the amount of VEGF bound to aflibercept. The mechanism of action of this drug suggested the use of TMDD model. Among several types, the MM approximation of the TMDD model developed by Gibiansky *et al.* [13] proved the most appropriate to reflect the kinetics of aflibercept in our study as free aflibercept concentrations were much higher than target concentrations and their binding to VEGF resulted in a fully saturated target. The relatively low concentration of free circulating VEGF was confirmed by the range of 19–47 pg ml^{-1} for plasma VEGF concentrations observed in healthy subjects in a meta-analysis [17]. In order to model simultaneously both free and bound drug, the MM equation of bound aflibercept was added into the system of differential equations. Free aflibercept is therefore eliminated through two pathways: non saturable elimination from the central compartment (k_{el}) and a specific and saturable binding to VEGF, followed by internalization (k_{int}) of bound aflibercept, which is its dominant elimination pathway. In addition, the formation

of bound aflibercept could occur predominantly in central or peripheral compartment. The final PK model was an approximate TMDD model involving two compartments for free aflibercept and one compartment for bound aflibercept, with MM type binding of free aflibercept to VEGF from the peripheral compartment.

The first-order dissociation rate constant (k_{off}) of the complex to give back free aflibercept was assumed to be negligible. Bound aflibercept might dominantly eliminate by internalization ($k_{\text{int}} > k_{\text{off}}$). This assumption is reasonable and consistent with the study of Eppler *et al.* on the development of a TMDD model for recombinant human VEGF (rhVEGF) after i.v. administration in patients with coronary artery disease [18]. VEGF binding to its high affinity receptors was concluded as an essentially irreversible process *in vivo*. Aflibercept might have the same properties because this drug is produced from two high affinity receptors of aflibercept (VEGFR1 and 2).

Recently, a new derivation of MM approximation of TMDD, called the irreversible binding MM model has been proposed by Gibiansky *et al.* when the dissociation rate constant is negligible and the free target concentration is much smaller than the free drug concentration [19]. The developed model of aflibercept is therefore close to this new approximation of the TMDD model. The only difference is that an extra differential equation was added to describe the evolution of bound drug. The MM parameters in the final model represent the combinations of the TMDD model parameters: $V_{\text{max}} = k_{\text{syn}} \cdot V_R$ and $K_m = K_{\text{IB}} = k_{\text{deg}}/k_{\text{on}}$, where V_R is the volume of distribution of target and K_{IB} , k_{syn} , k_{deg} are the irreversible binding constant, target production rate and target degradation rate, respectively. Although these estimates are not sufficient for the complete description of the full TMDD model, they can provide useful information about the underlying kinetics of drug and target.

During the model development, the general TMDD model with k_{on} and k_{off} was also implemented, assuming that the total target concentration was constant. Free VEGF was then the difference between total VEGF and VEGF binding to aflibercept. This model showed that k_{on} and k_{off} could not be estimated separately with very small value of k_{off} and its poor precision. Removing k_{off} resulted in the model developed by Eppler *et al.* [18], but we obtained unsatisfactory fits for bound aflibercept, either for central or peripheral binding to VEGF. A plateau phase after reaching the peak, instead of a decreasing phase was predicted for bound aflibercept and the predicted peak was lower.

Alternatively, we also evaluated the linear central or peripheral binding models with a first-order association rate constant (k_{on}) instead of V_{max} and K_m . These models did not describe the observed concentrations of bound aflibercept well, specifically the decreasing phase which illustrated the elimination of bound aflibercept after reaching the peak was poorly approached. They were also found less appropriate by statistical criteria.

The nonlinear peripheral binding model was found to describe better the kinetics of drugs than the central binding one, which was consistent with the previous studies on the distribution of VEGF in the body [17]. Large quantities of VEGF were reported in the extravascular space of tumour and skeletal muscle [17], suggesting an important source of endogenous VEGF in the peripheral compartment and supporting our choice of final model. The plasma concentrations of bound aflibercept in the data were assumed to be the same as those in the peripheral compartment, with rapid transfer between extravascular and plasma space.

In the final model, the clearance of bound aflibercept was found to be 6.3 times lower than that of free aflibercept from the central compartment (0.14 l day^{-1} and 0.88 l day^{-1} , respectively). Both free and bound aflibercept were eliminated quite slowly. The typical central volume of distribution of free aflibercept was 4.94 l , indicating that free aflibercept has a low level of tissue diffusion and that it circulates mostly in the extravascular spaces. The volume of distribution of bound aflibercept was close to that of free aflibercept, which was observed in intermediate models and was fixed to the mean value of V_p in the final model due to the problem of identifiability. In the prior non-compartmental analysis of free aflibercept, the average clearance was 0.97 l day^{-1} and the steady-state volume of distribution was 5.98 l . These values were similar to those estimated by the modelling approach. Compared with a similar anti-angiogenic agent, bevacizumab, the central volume of distribution was close but the clearance of total aflibercept was nearly four times faster than that of total bevacizumab [20]. From MM parameters, the production rate of VEGF and the irreversible binding constant were calculated to be 0.99 mg day^{-1} and $2.91 \mu\text{g ml}^{-1}$ or 19909 pM , respectively. The irreversible binding constant was much larger than the *in vitro* binding affinity K_d observed for VEGF-A. Such a difference is due to the negligible value of k_{off} *in vivo* and it means that the target degradation rate (k_{deg}) is significantly faster than the dissociation rate (k_{off}).

The advantage of this study is the availability of drug concentration data with free and bound forms since they were separately assayed, which is not the case for many antibody products. Generally, only total drug, representing the sum of free and bound was assayed [20–22]. The pharmacokinetic analysis of total drug could not provide a good explanation of the mechanism of action of the drug, including the binding to the target; neither was it able to characterize well its pharmacokinetic profile. For example, a simple model of two-compartment infusion with first-order elimination was published for bevacizumab and squalamine, two anti-angiogenic agents dosed in total form [20, 21]. The availability of complex data for aflibercept helped us to characterize both the linear and non linear elimination pathways, as well as the complex internalization, and estimate the mechanistic parameters of the

TMDD system. This could not be achieved using a classical MM model of free drug alone. The concentrations of bound aflibercept could be considered as pharmacodynamic data information of this drug. Our model development was in agreement with the guidelines of modelling for drugs with TMDD properties as mentioned by Yan *et al.* [23]. To our knowledge, the model that we developed is the first mechanism-based population pharmacokinetic model for an anti-VEGF drug.

In conclusion, the present PK model for aflibercept characterizes well the underlying mechanism of disposition of aflibercept, where a saturable, high-affinity binding of aflibercept to its pharmacologic target (VEGF) is responsible for the observed nonlinear pharmacokinetic behaviour of the free drug. Although further studies will be needed to assess the influence of covariates because of the limited size of the present sample, this model helps to understand the properties of aflibercept better and provides a useful support for further studies in patients during the clinical development, in particular the determination of therapeutic doses using bound aflibercept concentrations as a marker of VEGF inhibition.

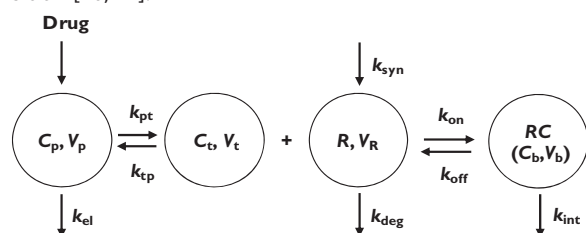
Competing Interests

Hoai-Thu Thai received a research grant from Sanofi-aventis to perform this research and is now doing a PhD funded by a research grant from Sanofi-aventis. Emmanuelle Comets acted as a consultant for Sanofi-aventis in 2009–10. France Mentré is the director of Inserm UMR 738 which receives a research grant from Sanofi-aventis and has been paid for running an education program at Sanofi-Aventis. Christine Veyrat-Follet and Gerard Sanderink are employees of Sanofi-aventis. Nicole Vivier and Catherine Dubruc are former employees of Sanofi-aventis.

The authors thank Global Metabolism and the Pharmacokinetics Department, Sanofi-aventis, Paris which supported Hoai-Thu Thai by a research grant during this work.

APPENDIX: Model development

General peripheral model of target-mediated drug disposition [10, 11]:



$$\begin{aligned}\frac{dC_p}{dt} &= \text{Input} - (k_{el} + k_{pt}) \cdot C_p + k_{tp} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{pt} \cdot \frac{C_p \cdot V_p}{V_t} - k_{tp} \cdot C_t - \frac{k_{on} \cdot C_t \cdot R \cdot V_R}{V_t} + k_{off} \cdot \frac{C_b \cdot V_b}{V_t} \\ \frac{dC_b}{dt} &= \frac{k_{on} \cdot C_t \cdot R \cdot V_R}{V_b} - k_{int} \cdot C_b - k_{off} \cdot C_b \\ \frac{dR}{dt} &= k_{syn} - k_{deg} \cdot R - k_{on} \cdot C_t \cdot R + k_{off} \cdot \frac{C_b \cdot V_b}{V_R}\end{aligned}\quad (A1)$$

where C_p, C_t, C_b, R are the concentrations of free drug in the central and peripheral compartment, complex (bound drug) and target, V_p, V_t are the volume of distribution of free drug in the central and peripheral compartment and V_R, V_b are the volume of distribution of target and bound drug in the peripheral compartment. k_{el} is the elimination rate of free drug from central compartment, k_{on}, k_{off} are the association and dissociation rate constants, k_{int} is the internalization rate constant of the complex and k_{syn}, k_{deg} are the target production and degradation rates.

To write this general model, we did not assume that the drug, the target and the bound drug have the same volume of distribution, and we assumed that the amount of bound drug created during a unit of time, produced in the compartment of distribution of the target, is $k_{on} \cdot R \cdot C_t \cdot V_R$ [10, 11].

When the free drug concentrations are much higher than target concentrations ($C \gg R$) and their binding results in a fully saturated target, the full TMDD can be replaced by the MM approximations of the TMDD model following the suggestions of Gibiansky *et al.* [13, 19]. We explain these approximations for peripheral TMDD models.

Approximation 1

The first MM approximation of TMDD model was proposed in case of reversible binding, based on the assumptions that the drug–target complex is in a quasi-steady-state or the derivative of the complex concentration, dC_b/dt , is zero [13]. V_R and V_b were assumed to be equal to V_t .

$$k_{on} \cdot C_t \cdot R - (k_{int} + k_{off}) \cdot C_b = 0 \quad (A2)$$

$$\frac{C_t \cdot R}{C_b} = \frac{k_{int} + k_{off}}{k_{on}} = K_{ss} \quad (A3)$$

$$C_b = \frac{R_{tot} \cdot C_t}{K_{ss} + C_t} \quad (A4)$$

where R_{tot} is the total concentration of target ($R_{tot} = R + C_b$).

We can rearrange the equation for dC_t/dt , by expressing C_b as a function of C_t and using the expression of K_{ss} , yielding the following equation:

$$\frac{dC_t}{dt} = k_{pt} \cdot \frac{C_p \cdot V_p}{V_t} - k_{tp} \cdot C_t - \frac{k_{int} \cdot R_{tot} \cdot C_t}{K_{ss} + C_t} \quad (A5)$$

This new equation dC_t/dt therefore includes a MM type elimination with

$$V_{\max} = R_{\text{tot}} \cdot k_{\text{int}}; K_m = K_{ss} = (k_{\text{int}} + k_{\text{off}})/k_{\text{on}} \quad (\text{A6})$$

If, in addition, the total target concentration R_{tot} can be considered constant, V_{\max} is a constant parameter of the system. The TMDD equation system then results in the two following equations:

$$\begin{aligned} \frac{dC_p}{dt} &= \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{V_{\max} \cdot C_t}{K_m + C_t} \end{aligned} \quad (\text{A7})$$

Approximation 2

When the binding is irreversible, the dissociation binding rate constant $k_{\text{off}} = 0$ then the TMDD equations simplify:

$$\begin{aligned} \frac{dC_p}{dt} &= \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{k_{\text{on}} \cdot C_t \cdot R \cdot V_R}{V_t} \\ \frac{dC_b}{dt} &= \frac{k_{\text{on}} \cdot C_t \cdot R \cdot V_R}{V_b} - k_{\text{int}} \cdot C_b \\ \frac{dR}{dt} &= k_{\text{syn}} - k_{\text{deg}} \cdot R - k_{\text{on}} \cdot C_t \cdot R \end{aligned} \quad (\text{A8})$$

The second MM approximation of the TMDD models was proposed in the case of irreversible binding. It is based on the assumption that the target is in a quasi steady-state so that the derivative of target concentration, dR/dt , is zero [19].

$$k_{\text{syn}} - k_{\text{deg}} \cdot R - k_{\text{on}} \cdot C_t \cdot R = 0 \quad (\text{A9})$$

$$R = \frac{k_{\text{syn}}}{k_{\text{deg}} + k_{\text{on}} \cdot C_t} = \frac{R_0 \cdot K_{\text{IB}}}{K_{\text{IB}} + C_t} \quad (\text{A10})$$

where K_{IB} is the irreversible binding constant ($K_{\text{IB}} = k_{\text{deg}}/k_{\text{on}}$) and R_0 is the target concentration at baseline ($R_0 = k_{\text{syn}}/k_{\text{deg}}$), which can be obtained assuming steady-state in the differential equation for R in the absence of drug.

Using this expression for R and noticing that $R_0 \cdot K_{\text{IB}} \cdot k_{\text{on}} = k_{\text{syn}}$, we can simplify the equation for dC_t/dt to:

$$\frac{dC_t}{dt} = k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{1}{V_t} \cdot \frac{k_{\text{syn}} \cdot V_R \cdot C_t}{K_{\text{IB}} + C_t} \quad (\text{A11})$$

Also, generally, only free drug is measured, so that in their model, Gibiansky *et al.* did not take into account the evolution of the complex (dC_b/dt) [19]. The TMDD equation system therefore includes only two equations:

$$\begin{aligned} \frac{dC_p}{dt} &= \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{1}{V_t} \cdot \frac{k_{\text{syn}} \cdot V_R \cdot C_t}{K_{\text{IB}} + C_t} \end{aligned} \quad (\text{A12})$$

Application to our study

In our case, the free aflibercept concentrations were much higher than the target concentrations. We also considered the evolution of bound aflibercept since the concentrations of bound aflibercept were available and could not be considered negligible in the system. In addition, the dissociation rate constant was found to be very small during model development so that irreversible binding and the negligible change in target concentrations (dR/dt) could be assumed, which is similar to the assumptions of the second MM approximation described above. Thus, our differential equation system is (A12) to which we add the equation for dC_b/dt from (A8) where $k_{\text{off}} = 0$.

Substituting the expression for $R = \frac{R_0 \cdot K_{\text{IB}}}{K_{\text{IB}} + C_t}$ derived from the second assumption of TMDD model, we find the following expression:

$$\begin{aligned} \frac{dC_p}{dt} &= \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{1}{V_t} \cdot \frac{k_{\text{syn}} \cdot V_R \cdot C_t}{K_{\text{IB}} + C_t} \\ \frac{dC_b}{dt} &= \frac{1}{V_b} \cdot \frac{k_{\text{syn}} \cdot V_R \cdot C_t}{K_{\text{IB}} + C_t} - k_{\text{int}} \cdot C_b \end{aligned} \quad (\text{A13})$$

Again, we notice a MM elimination for equation dC_t/dt , which enters the equation of dC_b/dt as a saturable input.

This equation system can be written with MM parameters, which represents our final model:

$$\begin{aligned} \frac{dC_p}{dt} &= \text{Input} - (k_{\text{el}} + k_{\text{pt}}) \cdot C_p + k_{\text{tp}} \cdot \frac{C_t \cdot V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{\text{pt}} \cdot \frac{C_p \cdot V_p}{V_t} - k_{\text{tp}} \cdot C_t - \frac{1}{V_t} \cdot \frac{V_{\max} \cdot C_t}{K_m + C_t} \\ \frac{dC_b}{dt} &= \frac{1}{V_b} \cdot \frac{V_{\max} \cdot C_t}{K_m + C_t} - k_{\text{int}} \cdot C_b \end{aligned} \quad (\text{A14})$$

where $V_{\max} = k_{\text{syn}} \cdot V_R = A_{\text{syn}}$ and $K_m = K_{\text{IB}}$.

For parameter estimation, the micro constants (k_{el} , k_{pt} , k_{tp}) can be replaced by the macro constants: $Q = k_{\text{tp}} \cdot V_t = k_{\text{pt}} \cdot V_p$ and $\text{CL} = k_{\text{el}} \cdot V_p$, where CL is the clearance of free drug from the central compartment and Q is the intercompartment clearance of free drug.

REFERENCES

- 1 Baka S, Clamp AR, Jayson GC. A review of the latest clinical compounds to inhibit VEGF in pathological angiogenesis. Expert Opin Ther Targets 2006; 10: 867–76.

- 2** Folkman J, Merler E, Abernathy C, Williams G. Isolation of a tumor factor responsible for angiogenesis. *J Exp Med* 1971; 133: 275–88.
- 3** Holash J, Davis S, Papadopoulos N, Croll SD, Ho L, Russell M, Boland P, Leidich R, Hylton D, Burova E, Ioffe E, Huang T, Radziejewski C, Bailey K, Fandl JP, Daly T, Wiegand SJ, Yancopoulos GD, Rudge JS. VEGF-Trap: a VEGF blocker with potent antitumor effects. *Proc Natl Acad Sci U S A* 2002; 99: 11393–8.
- 4** Moreira IS, Fernandes PA, Ramos MJ. Vascular endothelial growth factor (VEGF) inhibition – a critical review. *Anticancer Agents Med Chem* 2007; 7: 223–45.
- 5** Salven P, Manpaa H, Orpana A, Alitalo K, Joensuu H. Serum vascular endothelial growth factor is often elevated in disseminated cancer. *Clin Cancer Res* 1997; 3: 647–51.
- 6** Ferrara N. Vascular endothelial growth factor: basic science and clinical progress. *Endocrinol Rev* 2004; 25: 581–611.
- 7** Chu QS. Aflibercept (AVE0005): an alternative strategy for inhibiting tumour angiogenesis by vascular endothelial growth factors. *Expert Opin Biol Ther* 2009; 9: 263–71.
- 8** Lockhart AC, Rothenberg ML, Dupont J, Cooper W, Chevalier P, Sternas L, Buzenet G, Koehler E, Sosman JA, Schwartz LH, Gultekin DH, Koutcher JA, Donnelly EF, Andal R, Dancy I, Spriggs DR, Tew WP. Phase I study of intravenous vascular endothelial growth factor Trap, aflibercept, in patients with advanced solid tumors. *J Clin Oncol* 2009; 28: 207–14.
- 9** Sternberg CN. Systemic chemotherapy and new experimental approaches in the treatment of metastatic prostate cancer. *Ann Oncol* 2008; 19 (Suppl. 7): vii91–5.
- 10** Mager DE, Jusko WJ. General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J Pharmacokinet Pharmacodyn* 2001; 28: 507–32.
- 11** Mager DE. Target-mediated drug disposition and dynamics. *Biochem Pharmacol* 2006; 72: 1–10.
- 12** Mager DE, Krzyzanski W. Quasi-equilibrium pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *Pharm Res* 2005; 22: 1589–96.
- 13** Gibiansky L, Gibiansky E, Kakkar T, Ma P. Approximations of the target-mediated drug disposition model and identifiability of model parameters. *J Pharmacokinet Pharmacodyn* 2008; 35: 573–91.
- 14** Samson A, Lavielle M, Mentre F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Comput Stat Data Anal* 2006; 51: 1562–74.
- 15** Bertrand J, Comets E, Mentre F. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *J Biopharm Stat* 2008; 18: 1084–102.
- 16** Brendel K, Comets E, Laffont C, Laveille C, Mentre F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm Res* 2006; 23: 2036–49.
- 17** Kut C, Mac Gabhann F, Popel AS. Where is VEGF in the body? A meta-analysis of VEGF distribution in cancer. *Br J Cancer* 2007; 97: 978–85.
- 18** Eppler SM, Combs DL, Henry TD, Lopez JJ, Ellis SG, Yi JH, Annex BH, McCluskey ER, Zioncheck TF. A target-mediated model to describe the pharmacokinetics and hemodynamic effects of recombinant human vascular endothelial growth factor in humans. *Clin Pharmacol Ther* 2002; 72: 20–32.
- 19** Gibiansky L, Gibiansky E. Target-mediated drug disposition: new derivation of the Michaelis-Menten model, and why it is often sufficient for description of drugs with TMDD. 2010. PAGE 19; Abstr 1728. Available at <http://www.page-meeting.org/?abstract=1728> (last accessed 14 June 2011).
- 20** Lu JF, Bruno R, Eppler S, Novotny W, Lum B, Gaudreault J. Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol* 2008; 62: 779–86.
- 21** Bhargava P, Marshall JL, Dahut W, Rizvi N, Trocky N, Williams JI, Hait H, Song S, Holroyd KJ, Hawkins MJ. A phase I pharmacokinetic study of squalamine, a novel antiangiogenic agent, in patients with advanced cancers. *Clin Cancer Res* 2001; 7: 3912–9.
- 22** Hayashi N, Tsukamoto Y, Sallas WM, Lowe PJ. A mechanism-based binding model for the population pharmacokinetics and pharmacodynamics of omalizumab. *Br J Clin Pharmacol* 2007; 63: 548–61.
- 23** Yan X, Mager DE, Krzyzanski W. Selection between Michaelis-Menten and target-mediated drug disposition pharmacokinetic models. *J Pharmacokinet Pharmacodyn* 2010; 37: 25–47.

2.2 Analyse de pharmacocinétique de population d'affibercept chez les patients atteints de cancer

2.2.1 Résumé

Dans l'analyse précédente, la PK de l'affibercept chez les volontaires sains a été bien caractérisée par l'approximation MM du modèle TMDD avec une liaison irréversible de l'affibercept libre au VEGF dans le compartiment périphérique (THAI et al., 2011). Nous avons voulu appliquer ce modèle aux données de l'affibercept libre et lié chez les patients atteints de tumeurs solides afin de caractériser leur PK de population. Nous avons également examiné l'influence des divers facteurs physiopathologiques et des chimiothérapies sur leur PK et évalué les schémas posologiques proposés par simulation.

Les données de 9 essais cliniques (3 essais de phase I, 3 essais de phase II et 3 essais de phase III) avec 1506 patients atteints de cancer recevant l'affibercept en tant que monothérapie ou en association avec des chimiothérapies différentes ont été incluses dans cette analyse. L'affibercept a été administré par perfusion intraveineuse à des doses allant de 2 à 7 mg/kg toutes les 2 ou 3 semaines et jusqu'à 9 mg/kg toutes les 3 semaines. Nous avons utilisé l'algorithme SAEM implémenté dans le logiciel MONOLIX 4.1.2 pour estimer les paramètres du modèle. Dans un premier temps, nous avons construit le modèle de base et ensuite construit le modèle avec les covariables en utilisant le test de Wald. La sélection de covariables d'intérêt a été effectuée par la modélisation au lieu de l'utilisation des estimations bayésiennes des paramètres individuelles. Par la suite, nous avons utilisé le modèle final pour simuler les profils des concentrations de l'affibercept libre et lié pour plusieurs schémas posologiques différents.

Le modèle développé chez les volontaires sains décrit adéquatement la PK de l'affibercept chez les patients. Les paramètres obtenus chez les patients et chez les volontaires sains sont comparables sauf ceux caractérisant la fixation de l'affibercept au VEGF. La clairance et le volume de distribution de l'affibercept libre augmentent avec le poids corporel et ils sont plus faibles chez les femmes que chez les hommes. Les patients avec un taux d'albumine (ALB) faible ou d'alkaline phosphatase (ALK) élevée éliminent l'affibercept plus rapidement qu'un patient ayant des taux standards. La relation entre l'albumine et la clairance pourrait être expliquée par l'interaction entre le récepteur Fc néonatal (FcRn), de l'immunoglobuline (IgG) et de l'albumine (KUO et AVESON, 2011). L'étude de KIM et al. (2007) a montré que le FcRn recycle l'albumine, l'IgG endogène, et des produits biologiques à la base de l'IgG à travers les membranes cellulaires vers la circulation sanguine. Ainsi, de faibles niveaux d'albumine pourrait refléter une diminution du nombre de FcRn, et peut être associée à l'élimination plus rapide de l'affibercept. Notre analyse a également montré que le cancer du pancréas influence la fixation de l'affibercept au VEGF. Des simulations de différentes posologies ont montré que la saturation suffisante

du VEGF circulant dans l'intervalle posologique a été atteinte avec une dose de 4 mg/kg toutes les 2 semaines.

L'ensemble de ce travail est décrit dans un article soumis dans la revue *Cancer chemotherapy and pharmacology*.

2.2.2 Article 2 (accepté)

Population pharmacokinetic analysis of free and bound aflibercept in patients with advanced solid tumors

Hoai-Thu Thai^{1,2}, Christine Veyrat-Follet¹, France Mentré², Emmanuelle Comets²

¹ Drug Disposition Department, Sanofi, Paris, France

² INSERM, UMR 738, F-75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, UMR 738, F-75018 Paris, France

Correspondance:

Hoai-Thu Thai

Sanofi R&D
Disposition Safety & Animal Research
1, avenue Pierre Brossolette
91380 Chilly-Mazarin
France
Tel: + 33 1 60 49 65 12
Fax: +33 1 60 49 71 61
Email: hoai-thu.thai@sanofi.com

Keywords: Population pharmacokinetics, target-mediated drug disposition, aflibercept, VEGF, covariate effect, dosing schedule

Word count: 5100

Number of tables: 4

Number of figures: 6

Abstract

Objective Aflibercept (Zaltrap®) is a novel anti-angiogenic agent that binds to vascular endothelial growth factor (VEGF) and inhibits VEGF-dependent tumor growth. We aimed to characterize the population pharmacokinetics (PK) of free and bound aflibercept in patients with solid tumors, to examine the influence of covariates on their PK and to evaluate the proposed dosing regimens by simulation.

Methods Data from 9 clinical trials with 1506 cancer patients receiving aflibercept (2-9 mg/kg every 2 or 3 weeks; 1 hour IV infusion) as a monotherapy or in combination with various chemotherapies were included. Free and bound aflibercept concentrations were analyzed using a nonlinear mixed-effects modeling approach with MONOLIX 4.1.2.

Results An approximation of a target mediated-drug disposition model with irreversible binding of free aflibercept to VEGF adequately described the PK of free and bound aflibercept. The typical estimated clearances for free (CL_f) and bound aflibercept (CL_b) were 0.88 and 0.19 L/day, respectively. The volumes of distribution for free (V_p) and bound (V_b) aflibercept were similar (~4 L). CL_f and V_p increased with body weight and were lower in women. Patients with low albumin (ALB) or high alkaline phosphatase (ALK) had faster CL_f compared to a typical patient. Pancreatic cancer may be associated with changes in binding of aflibercept to VEGF. Simulations of different dosing regimens showed that adequate saturation of circulating VEGF was achieved with a dose of 4 mg/kg every 2weeks.

Conclusions Aflibercept kinetics was most affected by gender, body weight, ALB, ALK and pancreatic cancer. Simulations supported the rationale for the recommended dose of 4 mg/kg every 2 weeks for aflibercept.

Introduction

Malignant tumors are dependent on angiogenesis to maintain a source of nutrition and oxygen supporting their growth and metastasis [1]. Vascular endothelial growth factor (VEGF) is a homodimeric protein that activates two high-affinity receptors on the vascular endothelium (VEGFR1 and VEGFR2), promoting the formation of blood vessels that are required for normal and neoplastic tissue growth. VEGF has become a major target for anti-angiogenic therapy in managing cancers because its overexpression in several tumor types is associated with increased tumor vascularity, proliferation, progression, invasion, metastasis, and poor prognosis [2-4]. Preclinical studies have shown that inhibition of VEGF results in tumor growth inhibition in several human tumor xenografts in the nude mouse model, including colorectal cell lines [5].

Aflibercept (known as ziv-aflibercept in the United States; Sanofi, Paris, France, and Regeneron Pharmaceuticals, Tarrytown, NY) is a novel antiangiogenic agent [6]. It is a recombinant protein consisting of human VEGF receptor extracellular domains (domain 2 from VEGFR1 and domain 3 from VEGFR2) fused to the Fc portion of human immunoglobulin G1 (IgG1). Aflibercept binds to all isoforms of VEGF-A, VEGF-B and to placental growth factor (PlGF) [7]. It interferes with the biological actions of VEGF by forming a complex with VEGF in the blood stream and extravascular space and preventing it from interacting with its receptors on endothelial cells. The affinity of aflibercept for VEGF-A (K_d *in vitro* = 0.5 pM) is higher than that of anti-VEGF monoclonal antibodies, such as bevacizumab (K_d *in vitro* = 500 pM) [8,9].

Preclinical studies in mice have demonstrated the effects of aflibercept on capillary regression, narrowing of vessels, blood flow cessation as well as endothelial cell apoptosis [10]. Inhibition of tumor growth and tumor angiogenesis, inhibition of metastases and improved survival have been observed with aflibercept in tumor xenografts for various cancers [6,11,12]. Phase I dose-escalation studies, Phase II and Phase III trials have explored the antitumor activity of aflibercept as a single agent or in combination with a number of chemotherapy agents/regimens, including in patients with non-small cell lung, ovarian, pancreatic and colorectal cancers [6,13-16]. Recently, aflibercept at 4 mg/kg every 2 weeks plus FOLFIRI (combination of 5-fluorouracil, leucovorin and irinotecan) has been demonstrated to improve overall survival, progression-free survival and response rate in patients with metastatic colorectal cancer [17-19].

During clinical development, data is accrued from different sources, such as phase I trials in healthy subjects and phase I to III trials in cancer patients. Modelling pharmacokinetic (PK) data allows us to integrate data from different clinical trials to update our knowledge about aflibercept and support drug development strategy. Following this approach, a population PK study was initiated with the aim of characterising the PK of aflibercept and quantifying parameter changes in different populations.

Previously, we reported the results of a population PK analysis of free and VEGF-bound aflibercept in 56 healthy subjects receiving single intravenous (i.v.) doses of 1 to 4 mg/kg recruited in two phase I clinical studies [20]. An approximation of a target mediated-drug disposition (TMDD) model with Michaelis-Menten (MM) type binding of free aflibercept to VEGF in the peripheral compartment was used to describe the PK profile of both free and bound aflibercept. This model adequately characterized the nonlinear binding of aflibercept to VEGF in this healthy subject population.

The objectives of this analysis were (i) to characterize the population PK of free and bound aflibercept in cancer patients by extending the model developed in healthy subjects; (ii) to investigate covariates that could influence their PK and (iii) to evaluate various dosing regimens by simulation.

Methods

Patient population and study design

A total of 9 clinical trials of aflibercept from Phase I to Phase III in 1506 patients with advanced solid tumors were included in the population analysis. Aflibercept was used as a single agent in 4 trials and in combination with cytotoxic chemotherapy regimens in 5 trials. A summary of clinical trials included in the analysis is given in Table 1. Phase I trials were conducted in patients with various types of solid tumors while Phase II and Phase III trials were conducted in patients with ovarian, pancreatic, lung and colorectal cancer. In these studies, aflibercept was administered as a 1 hour i.v. infusion at dose levels ranging from 2 to 7 mg/kg every 2 or 3 weeks and up to 9 mg/kg in every 3 weeks schedule only. Dose adjustments and/or cycle delays were

permitted in case of toxicity. An intensive PK sampling scheme was implemented in Phase I studies as presented in Table 1. For Phase II and Phase III trials, PK samples were collected pre-dose and at the end of aflibercept infusion on Day 1 (Cycle 1), then every odd cycle before treatment administration and at approximately 30 and 90 days after the last aflibercept treatment. All study protocols were in accordance with recommendations of the 18th World Health Congress (Helsinki, 1964) and complied with the laws and regulations, as well as any applicable guidelines, of the countries where the studies were conducted. All patients gave written informed consent.

Assay method

For all clinical trials, free aflibercept and bound aflibercept (VEGF:aflibercept complex) plasma concentrations were measured using enzyme-linked immunosorbent assay (ELISA) method. The anti-drug-antibodies (ADA) were also detected by a titer-based, bridging immunoassay to evaluate the potential effect of immunogenicity on the PK of aflibercept. Blood samples were collected in tubes (containing 1 mL of citrate buffer, sodium citrate, and 4.2 mg of citric acid) and were centrifuged at 2000 g for 15 minutes at room temperature. Plasma was stored at -80°C until analyzed.

In the assay of free aflibercept, human VEGF₁₆₅ initially adsorbed to the surface of a polystyrene solid support was used to capture free aflibercept in the samples and then bound to a mouse anti-human specific VEGFR-1 monoclonal antibody. After that, the immobilized murine monoclonal complex was bound to a goat antimouse IgG antibody conjugated to Horseradish Peroxidase for detection. The limit of quantification (LOQ) for free aflibercept in plasma was initially 31.3 ng/mL (for TED6115/TED6116), and then 15.6 ng/mL (for subsequent studies).

The assay of bound aflibercept was similar to that of free aflibercept, except for the use of the anti-human VEGF₁₆₅ antibody instead of human VEGF₁₆₅ as the capture reagent in the coated plate. The LOQ for bound aflibercept in plasma was 43.9 ng/mL.

Since bound aflibercept contains one molecule of endogenous VEGF and one molecule of aflibercept, bound aflibercept concentrations were expressed as free aflibercept equivalents for

PK analyses using 0.717, the ratio of molecular weights between free and bound aflibercept. The units of free and bound aflibercept concentrations were $\mu\text{g/mL}$ and $\mu\text{g.eq/mL}$, respectively.

Population pharmacokinetic analysis

The population PK analysis was performed using a nonlinear mixed-effects modeling approach with MONOLIX 4.1.2 [21] implementing the SAEM algorithm [22]. The model control files were written using MLXTRAN script. Graphical analyses for model evaluation were performed with R 2.14.1.

Structural model

The previous population PK analysis of free and bound aflibercept in healthy subjects demonstrated that the irreversible MM (IB-MM) approximation of TMDD model was the best PK model for aflibercept. This model is shown in Figure 1 and includes 2 compartments for free aflibercept, 1 compartment for bound aflibercept and irreversible MM type binding of free aflibercept to VEGF in the peripheral compartment [20]. In addition, the dissociation rate constant (k_{off}) which gives back free aflibercept and free VEGF was assumed to be negligible and is not represented on the Figure 1. The same model was applied to patient data.

In this model, the concentration of free aflibercept in central compartment (C_p), in tissue compartment (C_t) and the concentration of bound aflibercept (C_b) are described by the following system of differential equations:

$$\begin{aligned}\frac{dC_p}{dt} &= \frac{R_{inf}}{V_p} - (k_{el} + k_{pt}).C_p + k_{tp} \cdot \frac{C_t.V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{pt} \cdot \frac{C_p.V_p}{V_t} - k_{tp}.C_t - \frac{1}{V_t} \cdot \frac{V_{max}.C_t}{K_m + C_t} \\ \frac{dC_b}{dt} &= \frac{1}{V_b} \cdot \frac{V_{max}.C_t}{K_m + C_t} - k_{int}.C_b\end{aligned}$$

in which R_{inf} is the infusion rate, k_{el} (day^{-1}) is the first order elimination rate constant of free aflibercept from the central compartment, k_{tp} and k_{pt} (day^{-1}) are the first order rate constants between the central and the peripheral compartment, k_{int} (day^{-1}) is the first order rate constant of bound aflibercept internalization, V_p (L) is the central volume of distribution of free aflibercept,

V_t (L) is the peripheral volume of distribution of free aflibercept, V_b (L) is the volume of distribution of bound aflibercept, V_{\max} (mg/day) is the maximum binding capacity, K_m ($\mu\text{g/mL}$) is the concentration of free aflibercept corresponding to half of maximum binding capacity.

The clearances of free and bound aflibercept (CL_f and CL_b) and the intercompartmental clearance of free aflibercept (Q) between the central and the peripheral compartments are derived from the micro-constant as follows:

$$\begin{aligned} CL_f &= k_{el} \cdot V_p \\ CL_b &= k_{int} \cdot V_b \\ Q &= k_{tp} \cdot V_t = k_{pt} \cdot V_p \end{aligned}$$

The parameters to be estimated in this model are: CL_f , V_p , Q , V_t , CL_b , V_b , V_{\max} , and K_m .

Statistical model

Denoting f the function describing the PK structural model, the statistical model for observed concentration C_{ij} of subjects i for sampling time t_{ij} is:

$$C_{ij} = f(\theta_i, t_{ij}) + \varepsilon_{ij}$$

where θ_i is the vector of parameters of subject i and ε_{ij} is the residual error.

The interindividual variability (IIV) on all parameters was modeled with an exponential model, implying a log-normal distribution for the parameters, e.g. for CL :

$$CL_i = CL e^{\eta_{CL,i}}$$

where $\eta_{CL,i}$ denotes the random effect in subject i , CL_i the individual clearance parameter and CL the typical value of the population. Random effects were assumed to follow a normal distribution with zero mean and variance matrix Ω , which was modeled as diagonal.

The residual variability was modeled using a combined additive and proportional model for both free and bound aflibercept. The residual errors for free aflibercept ($\varepsilon_{free,ij}$) and bound aflibercept ($\varepsilon_{bound,ij}$) are assumed to be independent and normally distributed with mean zero and a heteroscedastic variance $\sigma^2_{free,ij}$ and $\sigma^2_{bound,ij}$ respectively, given by:

$$\sigma_{free,ij}^2 = (\sigma_{fa} + \sigma_{fp}f(\theta_i, t_{ij}))^2$$

$$\sigma_{bound,ij}^2 = (\sigma_{ba} + \sigma_{bp}f(\theta_i, t_{ij}))^2$$

where σ_{fa} and σ_{fp} are respectively the additive and proportional coefficients of the residual error model of free aflibercept; σ_{ba} and σ_{bp} are respectively the additive and proportional coefficients of the residual error model of bound aflibercept.

Estimation of potential interoccasion variability (IOV) was not performed in this analysis because the data was very sparse, containing a trough concentration every two cycles for 90% of the patients.

The variability models were investigated. For parameters with an estimated low level of IIV, we tested whether IIV could be removed from the model. The likelihood ratio test was used to discriminate between variability models through the difference in log likelihood (-2LL), computed using important sampling [23]. A p-value of 0.05 was considered statistically significant.

Covariate analysis

The covariate analysis was performed after obtaining the base model and focused on the main parameters describing the PK of free and bound aflibercept and their binding to VEGF: CL_f , V_p , V_{max} , K_m and CL_b .

The following covariates were included in the analysis: patient demographics (age, gender, weight and ethnicity), laboratory measurements at baseline (albumin (ALB), serum alkaline phosphatase (ALK), total bilirubin (BIL), aspartate amino transferase (AST), alanine amino transferase (ALT), total protein (TP), and creatinine clearance (CLCR)) and concomitant chemotherapy (irinotecan/5-FU/LV in TCD6118, docetaxel in VITAL and TDC6120, gemcitabine in VANILLA and FOLFIRI in VELOUR).

In this work, the effect of study and the effect of cancer type were not tested as these were confounded with the effect of chemotherapy (see Table 1). The effect of baseline endogenous

VEGF concentrations on PK aflibercept would have been interesting to investigate; however the large amount of missing values (62% of patients) precluded this as a covariate in the modeling.

The parameter-covariate relationships were modeled multiplicatively as follows (e.g. for CL):

$$CL_i = CL \left(\frac{COV_i}{COV_{median}} \right)^{\beta} e^{\eta_{CL,i}} \quad \text{for continuous covariates}$$

where β are the regression coefficient to be estimated, COV_{median} is the median value of covariates.

or $CL_i = CL \cdot e^{\beta \cdot COV_i} e^{\eta_{CL,i}}$ for dichotomous covariates COV_i taking 0 or 1 values

The construction of the covariate model was performed with hypothesis testing using the Wald test [24] in two steps:

- Step 1: For covariate screening, each potential parameter-covariate relationship was included one by one in the base model and the parameters were estimated. The significance of a covariate effect was then assessed using the Wald test.
- Step 2: All the significant covariates were included in the base model. The final model was built using a backward stepwise procedure by removing the non-significant covariates one by one, starting from the full model and removing the effect having the largest non-significant p-value of the Wald test. This step was repeated until only significant covariates remain in the model.

A p-value of 0.05 was considered statistically significant in both steps.

Model evaluation

Internal evaluation of the model was based on goodness-of-fit (GOF) plots, including plots of observations versus individual and population predictions. Plots of normalized prediction distribution error (NPDE) [25,26] versus time since last dose (TimeL) were used to assess model predictive performance, based on the simulation of 1000 datasets. The NPDE plots with 95% prediction intervals around the 10th, 50th and 90th percentiles were generated using the package NPDE 2.0 for R [27]. We also computed the η -shrinkage for each parameter and ε -shrinkage separately for free and bound aflibercept data to quantify the amount of information in the

individual data about the parameters [28]. Model evaluation was performed for both the base model and the final model.

Model-based simulations

In order to compare the time course of free and bound aflibercept concentrations after different dosing regimens (2, 4 or 6 mg/kg q2wk or q3wk), the estimates of the fixed and random effects obtained in the final model were used to simulate the steady-state PK profiles of free and VEGF-bound aflibercept in 1000 virtual patients. To mimic the actual treated population, the covariates were obtained by resampling 1000 patients from the 1506 patients in the study and collecting the set of covariates characterising each patient. The median bound concentrations of different doses for every 2 and 3 weeks regimens were compared to evaluate the optimal dosing regimen for achieving VEGF blockade. The median, 5th and 95th percentiles of free and bound aflibercept concentrations versus time were plotted for the recommend dose (4 mg/kg q2wk) to confirm the saturation of the binding of free aflibercept to VEGF during the treatment period.

Results

Demographics

The database for this population PK analysis consisted of 1506 patients including 151 patients from Phase I studies, 282 patients from Phase II studies and 1073 patients from Phase III studies. Patients found to be ADA positive were only found to have low titer levels and ADA positivity did not result in any observed impact on aflibercept PK (data not shown). As the result, all ADA positive patients (less than 5% of the analysis population) including those found to be ADA positive only at baseline were included in this analysis. The characteristics of these patients are shown in Table 2.

The majority of patients included in this analysis were Caucasians aged 65 years old and over and equally balanced between genders. Almost all the patients had normal renal function (CLCR: >80 mL/min) or mild renal dysfunction (CLCR: 50-80 mL/min). In this analysis, approximately one third of the patients had colorectal cancer, one third had lung cancer and the remaining third had other various solid tumors. Most patients received aflibercept in combination with chemotherapy.

Approximately 90% of patients were recruited in phase II and III studies and had limited pharmacokinetic sampling with only one peak and multiple trough concentrations of free and bound aflibercept.

Population pharmacokinetic model building

Base model development was performed using a data set including 1506 patients with 7916 free aflibercept concentrations and 6977 bound aflibercept concentrations greater than LOQ. The below quantification limit (BQL) data for free aflibercept (5.7%) and bound aflibercept (17.2%) were omitted in this analysis. The concentrations of both free and bound aflibercept were measured for 1378 patients while only free aflibercept concentrations were available for 128 remaining patients.

The TMDD model with IB-MM approximation previously developed in healthy subjects provided an adequate fit to the PK data of both free and bound aflibercept in patients. The IIV on maximum binding capacity (V_{max}) was found to be small (6%) and poorly estimated. Removing the variability for this parameter did not degrade the fit, with even a decrease of 5 points in -2LL values, probably as a result of improved model stability and estimation of other parameters. A combined additive and proportional residual error was retained for both free and bound aflibercept. The parameter estimates of the base model are shown in Table 3. Both fixed and random effects were precisely estimated with relative standard errors (RSEs) of less than 22%. The η -shrinkage was large for most parameters, reflecting the lack of information in the patients with sparse sampling, representing 90% of the subjects in the pooled dataset: CL_f (37%), V_p (69%), Q (86%), V_t (64%), V_b (73%), K_m (95%), CL_b (48%). The ε -shrinkage was 23 % for the data of free aflibercept and 19% for the data of bound aflibercept.

Seventeen potential covariates were evaluated for aflibercept by testing them on CL_f , V_p , K_m , and CL_b in the base model. They were not tested on V_{max} as its IIV was set to zero. Based on the Wald test of the univariate analysis using modeling, the following covariates were found to have an influence on PK parameters and were considered for inclusion in the model: gender, age, weight, CLCR, ALB, ALK, AST, ALT and all the concomitant chemotherapies. Ethnicity appeared to have no significant effect on aflibercept PK. After removing non significant covariates one by

one starting from the full model, the final model was achieved, with parameter estimates and the coefficients of all significant covariates presented in the Table 3.

The parameters in the final model were estimated with reasonable precision: RSEs $\leq 10\%$ for PK parameters, RSEs $\leq 42\%$ for random effects and RSEs $\leq 46\%$ for the coefficients of the significant covariates. The PK parameter estimates of the final model were similar to those of the base model. In the final model, the typical estimated clearance for free aflibercept was about 4.6 times faster than that of bound aflibercept (0.88 and 0.19 L/day, respectively). The volumes of distribution for free aflibercept (V_p) and bound aflibercept (V_b) were similar (~ 4 L). The maximum binding capacity was 0.82 mg/day and the concentration of free aflibercept corresponding to half of maximum binding capacity in this patient data set was 1.92 $\mu\text{g/mL}$.

The IIV on model parameters was moderate to high, ranging from 22.3% (CL_b) to 85.5% (CL_f). There was a small decrease in the estimates of the variabilities of all parameters, except Q and V_i when including covariates in the model. Residual variability was moderate for free aflibercept (proportional errors of 32.9%) but low for bound aflibercept (proportional errors of 9.01%).

The goodness-of-fit plots of the final model with covariates are shown in Figures 2 and 3. The plots of observations versus population and individual predictions (Figure 2) indicated that the model adequately described the observations despite an underprediction of high concentrations of free aflibercept. The plots of NPDE versus time since last dose (timeL) are presented in Figure 3 for the duration of the study (timeL ≤ 22 weeks). Here, NPDE plots are more appropriate than VPC (visual predictive check) plots because of the heterogeneity in sample times and doses [25,26]. An inset is shown plotting only the data for 6 weeks after the last treatment (timeL ≤ 6 weeks), which was the period of time where most of the observations were collected. The NPDE plots showed a symmetric distribution around zero for both free and bound aflibercept. The prediction bands indicated good model adequacy except for two issues. First, prediction intervals appear to be too large at timeL ≥ 14 weeks; this could be an artifact due to the small number of measurements late after the last dose. Second, the variability at early times for bound aflibercept is underestimated; this could be a consequence of the BQL data not being included in the database. Individual plots for free and bound aflibercept are shown in Figure 4 for 4 subjects from 4 different studies. For most subjects, the model described reasonably well the observations for both free and bound aflibercept.

Assessment of covariate effects

The magnitude of the effect of covariates on aflibercept PK parameters in the final model is shown in Table 4. Gender, body weight, ALB and ALK had the largest effects on CL_f . CL_f increased 12% for patients with a WT of 99 kg compared to the value of a typical 70 kg male subject and was 14% smaller in women compared to men. A patient with low serum ALB (normalized ALB of 0.57) would be expected to have 14.06% faster CL_f while patients with a high concentration of ALK (normalized ALK of 1.01) would have a 14.41% faster CL_f compared to a patient with median values of ALB and ALK. Gender and body weight also had significant effects on V_p with a 14.6% increase for a WT of 99 kg and a 19% decrease in women. Conversely, the covariate effects on CL_b were very small with less than 10% changes in CL_b compared to a typical subject, except for the effect of chemotherapy for irinotecan/5-FU/LV with a decrease of 12% of CL_b . The most important covariate associated with a change in the binding kinetics of aflibercept to VEGF, expressed by the irreversible binding constant K_m , was observed in patients with pancreatic cancer treated with gemcitabine in the study VANILLA. These patients had a K_m increased by 82% compared to the typical patient receiving aflibercept as a monotherapy. A decrease of 20% in K_m was also observed in women compared to typical men while an increase in K_m was also observed in patients with high level of ALT or low level of AST compared to the typical patient. However, the high positive correlation between ALT and AST implies a small overall effect on K_m , with less than 10% of change compared to the typical value of 1.92. Combining all the covariate effects on K_m explained 31.5 % of the IIV in this parameter. For other parameters, the covariate effects explained very little the IIV on their parameters: 12.2% for CL_f , 16.4% for V_p and 5% for CL_b .

Simulation of various dosing regimens

The final PK model was used to simulate the concentration-time courses at steady-state of free and bound aflibercept for 1000 virtual patients receiving 8 doses of 2, 4 or 6 mg/kg every 2 weeks or 6 doses of 2, 4 or 6 mg/kg every 3 weeks. Figure 5 presents the predicted median profiles of bound aflibercept at steady-state for these different dosing regimens. Similar bound aflibercept levels were observed between 4 mg/kg and 6 mg/kg for q2wk regimen, indicating saturation of binding of aflibercept to circulating VEGF was reached at doses ≥ 4 mg/kg. In addition, same bound aflibercept levels were shown between the 4 mg/kg q2wk regimen and the

6 mg/kg q3wk regimen, suggesting similar saturation of circulating VEGF for these two regimens. The predicted steady-state concentrations of free and bound aflibercept are presented in Figure 6. It illustrated that the free aflibercept concentrations remain greater than bound aflibercept concentrations throughout all the dosing intervals in most patients receiving 4 mg/kg q2wk. As aflibercept binds to VEGF with a 1:1 ratio, maintaining free drug concentrations above bound drug concentrations throughout the dosing intervals would maximize binding of aflibercept to endogenous VEGF.

Discussion

In this study, we present a population PK analysis of free and bound aflibercept in 1506 patients with advanced solid tumors from 9 clinical trials. Aflibercept was administered intravenously every two weeks as a single agent, every two or three weeks in combination with various chemotherapy drugs at dose levels ranging from 2 to 7 mg/kg and up to 9 mg/kg in the every 3 weeks schedule only. The influence of covariate effects on aflibercept pharmacokinetics was studied and quantified.

Based on the mechanism of action, aflibercept exhibits a target-mediated drug disposition (TMDD) as the binding of the drug to the target influence the pharmacokinetics of the drug. A general TMDD model describes the elimination pathway of drug plasma concentrations as the combination of first-order elimination from the central compartment and specific target binding clearance followed by internalization of drug-target complex [29]. The binding of the drug to the target can occur predominantly in the central or the peripheral compartment [30]. The full TMDD model is however complex and generally overparameterised. As a result, several approximations of TMDD model have been proposed: quasi equilibrium (QE), quasi steady state (QSS) and Michaelis-Menten (MM) [31-33]. The QE approximation is based on the assumption that the drug-target binding is much faster than all other system processes. The QSS approximation assumes the drug-target complex concentration changes more slowly than the binding and internalization process. The MM approximation describes the system when the target concentration is small relative to the free drug concentration and the dosing regimens result in the target being fully saturated [32]. Recently, a new derivation of MM approximation of TMDD model, the irreversible binding MM (IB-MM) model, can be used when the dissociation rate constant is negligible [33]. This is the model we used previously in healthy subjects. In this

model, the MM parameters (V_{max} , K_m) are related to TMDD model parameters: $V_{max}=k_{syn}.V_R$ and $K_m=K_{IB}=k_{deg}/k_{on}$, where V_R is the volume of distribution of target and K_{IB} , k_{syn} , k_{deg} are the irreversible binding constant, target production rate and target degradation rate, respectively [6].

In the present study, the same structural model was applied successfully in patients, suggesting the similarity in binding kinetics of aflibercept to VEGF in cancer patients and healthy subjects: irreversible binding occurring dominantly in the peripheral compartment (extravascular space). This is in agreement with large quantities of VEGF in tumors and skeletal muscle [34]. Compared to healthy subjects, the population estimates for the clearance of free aflibercept in typical 70 kg male cancer patients were the same (0.88 L/hr) while the clearance of bound aflibercept was slightly faster (0.14 vs 0.19 L/day). Similar values of V_p and V_b (around 4 L) were also observed, indicating a low level of tissue diffusion of aflibercept in both healthy subjects and patients. In this analysis, the volume of distribution V_b was correctly estimated with similar value to V_p while it had to be fixed to the value of V_p in the analysis of healthy subjects due to identifiability issues. The MM parameters (V_{max} and K_m), reflecting the binding of aflibercept to VEGF, were however lower in cancer patients than in healthy subjects (0.82 mg/day and 1.92 μ g/mL vs 0.99 mg/day and 2.91 μ g/mL, respectively). These findings were not expected. K_m should be similar for these two populations and V_{max} should be higher in cancer patients because of the faster secretion rate of VEGF and larger volume of distribution of VEGF in tumor tissue than in healthy tissue [35]. The comparison of binding kinetics between healthy subjects and patients is somewhat difficult because the designs and the studied doses were quite different in these two populations. The binding parameter estimates obtained in the previous study for healthy subjects may be impacted by the lower doses (e.g 1 mg/kg) given as single administration, and non available data of bound aflibercept at late time points compared to those measured in patients. It may also explain slightly faster clearance estimation for bound aflibercept. However, PK findings in healthy volunteers addressed many aspects of general clinical pharmacology and helped us to well identify the model structure thanks to homogenous data before moving to heterogeneous data pooled from different clinical studies, containing a lot of sparse data.

The covariate screening was performed using the modeling approach rather than using the Empirical Bayes Estimates of individual parameters because of the large shrinkage of parameters

in the base model. A backward stepwise elimination procedure was used to build the final model using the Wald test. Among all covariates tested, gender had the largest effects on both CL_f and V_p while gemcitabine had the largest effect on K_m . Body weight also had important effect on CL_f and V_p . The impact of the correlation between body weight and gender was found to be less significant due to the large variabilities in body weight observed in men and women. Moreover, the inclusion of both gender and body weight improved the model fit.

Patients with low serum albumin concentrations or high concentrations of alkaline phosphatase had approximately 14% faster CL_f and 32% lower exposure to free aflibercept compared to a typical patient with normal ALB or ALK. These effects were also found for total bevacizumab [36]. Low albumin and high alkaline phosphatase are generally indicative of disease severity and tumor burden [36] and may be the underlying cause of this association. In addition, an effect of albumin on clearance has been shown in the PK analysis of the monoclonal antibody, infliximab in patients with ulcerative colitis [37].

A finding of note in this study is the important increase (82%) of the irreversible binding constant (K_m or K_{IB}) found in pancreatic cancer patients treated with gemcitabine in the VANILLA study. However, it may be not be reasonable to assume that the affinity of aflibercept for VEGF varies from one cancer to another, and this increase may in fact reflect differences in V_{max} . Indeed, in the first steps of modeling, variability was included in the model for both V_{max} and K_m , but the former was removed because the model was not stable enough and the variability was very small and poorly estimated. However, since the correlation between V_{max} and K_m was very high (0.9), the apparent elevated value of K_m may indicate a decrease in maximal binding capacity (V_{max}) or in production rate of VEGF (k_{syn}). This would be related to the poorly vascularised nature of pancreatic tumors [38]. Unlike many other solid tumor types where the formation of new blood vessels (angiogenesis) promotes tumor growth by ensuring proper blood supply to the tumor, delivering nutrients and oxygen; pancreatic tumors do not need to create new blood vessels. They are able to survive with poor vasculature and under very low oxygen conditions, which makes drug delivery to pancreatic tumors especially difficult [39]. This has been suggested to be the reason why pancreatic cancer patients often have poor response to chemotherapy, even in combination with anti-VEGF agents administered intravenously [39,40].

Although several covariates were found significant in the final model, their contribution to aflibercept pharmacokinetics was mostly small with potential changes less than 20% (ranging from 3 to 18%) in CL_f , CL_b and V_p . As a result, the impact on VEGF inhibition would be minimal from a pharmacodynamic point of view. Conversely, the change in the maximal binding capacity of aflibercept to VEGF in patients with pancreatic cancer was much more important and may have contributed to the outcome of the VANILLA study in pancreatic cancer with the 4 mg/kg every 2 weeks regimen for aflibercept.

The final PK model was used to simulate the concentration-time courses of free and VEGF-bound aflibercept for 1000 virtual patients for different doses of aflibercept. Assuming the level of VEGF-bound aflibercept can be used as a marker of VEGF blockade [41], this simulation allowed us to confirm the choice of the recommended dose of 4 mg/kg every 2 weeks which is sufficient to saturate circulating VEGF in most patients. The developed model can be used to simulate and predict the concentration-time profiles of free and bound aflibercept in a patient population of interest, for example obese patients and patients with low albumin, with new dosing regimens. From a clinical perspective, it would be interesting to model the relationship between concentrations and a direct marker of aflibercept efficacy, such as the tumor sizes.

In summary, the present model adequately described the pharmacokinetics of free and bound aflibercept in cancer patients. The most important covariates affecting aflibercept kinetics were gender, body weight, ALB, ALK and pancreatic cancer. This model supported the rationale for the recommended dose of 4 mg/kg every 2 weeks for aflibercept in colorectal cancer.

Conflict of Interests

During this work, Hoai-Thu Thai was a PhD student with a research grant funded by Sanofi and UMR 738 INSERM- University Paris Diderot (head Professor France Mentré) had a research grant from Sanofi. Christine Veyrat-Follet is employee of Sanofi.

References

1. Folkman J (1997) Addressing tumor blood vessels. *Nat Biotechnol* 15 (6):510
2. Takahashi Y, Kitadai Y, Bucana CD, Cleary KR, Ellis LM (1995) Expression of vascular endothelial growth factor and its receptor, KDR, correlates with vascularity, metastasis, and proliferation of human colon cancer. *Cancer Res* 55 (18):3964-3968

3. Takebayashi Y, Aklyama S, Yamada K, Akiba S, Aikou T (1996) Angiogenesis as an unfavorable prognostic factor in human colorectal carcinoma. *Cancer* 78 (2):226-231
4. Radinsky R, Ellis LM (1996) Molecular determinants in the biology of liver metastasis. *Surg Oncol Clin North Am* 5 (2):215-229
5. Ferrara N, Davis-Smyth T (1997) The biology of vascular endothelial growth factor. *Endocr Rev* 18 (1):4-25
6. Gaya A, Tse V (2012) A preclinical and clinical review of aflibercept for the management of cancer. *Cancer Treat Rev* 38 (5):484-493
7. Sternberg CN (2008) Systemic chemotherapy and new experimental approaches in the treatment of metastatic prostate cancer. *Ann Oncol* 19 Suppl 7:vii91-95
8. Chu QS (2009) Aflibercept (AVE0005): an alternative strategy for inhibiting tumour angiogenesis by vascular endothelial growth factors. *Exp Opin Biol Ther* 9 (2):263-271
9. Lockhart AC, Rothenberg ML, Dupont J, Cooper W, Chevalier P, Sternas L, Buzenet G, Koehler E, Sosman JA, Schwartz LH, Gultekin DH, Koutcher JA, Donnelly EF, Andal R, Dancy I, Spriggs DR, Tew WP (2009) Phase I study of intravenous vascular endothelial growth factor trap, aflibercept, in patients with advanced solid tumors. *J Clin Oncol* 28 (2):207-214
10. Baffert F, Le T, Sennino B, Thurston G, Kuo CJ, Hu-Lowe D, McDonald DM (2006) Cellular changes in normal blood capillaries undergoing regression after inhibition of VEGF signaling. *Am J Physiol* 290 (2):H547-559
11. Holash J, Davis S, Papadopoulos N, Croll SD, Ho L, Russell M, Boland P, Leidich R, Hylton D, Burova E, Ioffe E, Huang T, Radziejewski C, Bailey K, Fandl JP, Daly T, Wiegand SJ, Yancopoulos GD, Rudge JS (2002) VEGF-Trap: a VEGF blocker with potent antitumor effects. *Proc Nat Acad Sci USA* 99 (17):11393-11398
12. Eichten A, Adler A, Cooper B, Griffith J, Wei Y, Yancopoulos G, Lin H, Thurston G (2012) Rapid decrease in tumor perfusion following VEGF blockade predicts long-term tumor growth inhibition in preclinical tumor models. *Angiogenesis* 16(2):429-441
13. Freyer G, Isambert N, You B, Zanetta S, Falandry C, Favier L, Trillet-Lenoir V, Assadourian S, Soussan-Lazard K, Ziti-Ljajic S, Fumoleau P (2012) Phase I dose-escalation study of aflibercept in combination with docetaxel and cisplatin in patients with advanced solid tumours. *British J Cancer* 107(4):598-603
14. Coleman RL, Duska LR, Ramirez PT, Heymach JV, Kamat AA, Modesitt SC, Schmeler KM, Iyer RB, Garcia ME, Miller DL, Jackson EF, Ng CS, Kundra V, Jaffe R, Sood AK (2011) Phase 1-2 study of docetaxel plus aflibercept in patients with recurrent ovarian, primary peritoneal, or fallopian tube cancer. *Lancet Oncol* 12 (12):1109-1117
15. Colombo N, Mangili G, Mammoliti S, Kalling M, Tholander B, Sternas L, Buzenet G, Chamberlain D (2012) A phase II study of aflibercept in patients with advanced epithelial ovarian cancer and symptomatic malignant ascites. *Gynecol Oncol* 125 (1):42-47
16. Ramlau R, Gorbunova V, Ciuleanu TE, Novello S, Ozguroglu M, Goksel T, Baldotto C, Bennouna J, Shepherd FA, Le-Guennec S, Rey A, Miller V, Thatcher N, Scagliotti G (2012) Aflibercept and docetaxel versus docetaxel alone after platinum failure in patients with advanced or metastatic non-small-cell lung cancer: a randomized, controlled phase III trial. *J Clin Oncol* 30 (29):3640-3647
17. He K, Cui B, Li G, Wang H, Jin K, Teng L (2012) The effect of anti-VEGF drugs (bevacizumab and aflibercept) on the survival of patients with metastatic colorectal cancer (mCRC). *OncoTargets and therapy* 5:59-65
18. Wang TF, Lockhart AC (2012) Aflibercept in the treatment of metastatic colorectal cancer. *Clin Med Insights* 6:19-30

19. Van Cutsem E, Tabernero J, Lakomy R, Prenen H, Prausova J, Macarulla T, Ruff P, van Hazel GA, Moiseyenko V, Ferry D, McKendrick J, Polikoff J, Tellier A, Castan R, Allegra C (2012) Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J Clin Oncol* 30 (28):3499-3506
20. Thai HT, Veyrat-Follet C, Vivier N, Dubruc C, Sanderink G, Mentre F, Comets E (2011) A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects. *British J Clin Pharmacol* 72 (3):402-414
21. Lavielle M (2008) MONOLIX (MOdèles NON LINéaires à effets miXtes). Orsay, France: MONOLIX group.
22. Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal* 49 (4):1020-1038
23. Samson A, Lavielle M, Mentre F (2007) The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat Med* 26 (27):4860-4875
24. Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans Amer Math Soc* 54 (3):426-482
25. Brendel K, Comets E, Laffont C, Laveille C, Mentre F (2006) Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm Res* 23 (9):2036-2049
26. Comets E, Brendel K, Mentre F (2008) Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput Methods Programs Biomed* 90 (2):154-166
27. Comets E, Brendel K, Mentre F (2010) Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics. *J Soc Fr Stat* 151:106-128
28. Savic RM, Karlsson MO (2009) Importance of shrinkage in empirical Bayes estimates for diagnostics: problems and solutions. *AAPS J* 11 (3):558-569
29. Mager DE, Jusko WJ (2001) General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J Pharmacokinet Pharmacodyn* 28 (6):507-532
30. Mager DE (2006) Target-mediated drug disposition and dynamics. *Biochem Pharmacol* 72 (1):1-10
31. Mager DE, Krzyzanski W (2005) Quasi-equilibrium pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *Pharm Res* 22 (10):1589-1596
32. Gibiansky L, Gibiansky E, Kakkar T, Ma P (2008) Approximations of the target-mediated drug disposition model and identifiability of model parameters. *J Pharmacokinet Pharmacodyn* 35 (5):573-591
33. Gibiansky L, Gibiansky E (2010) Target-Mediated Drug Disposition: New Derivation of the Michaelis-Menten Model, and Why It Is Often Sufficient for Description of Drugs with TMDD. *PAGE* 19, Abstr 1728 [www.page-meeting.org/?abstract=1728]
34. Kut C, Mac Gabhann F, Popel AS (2007) Where is VEGF in the body? A meta-analysis of VEGF distribution in cancer. *British J Cancer* 97 (7):978-985
35. Stefanini MO, Wu FT, Mac Gabhann F, Popel AS (2008) A compartment model of VEGF distribution in blood, healthy and diseased tissues. *BMC Syst Biol* 2:77
36. Lu JF, Bruno R, Eppler S, Novotny W, Lum B, Gaudreault J (2008) Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol* 62 (5):779-786

37. Fasanmade AA, Adedokun OJ, Ford J, Hernandez D, Johanns J, Hu C, Davis HM, Zhou H (2009) Population pharmacokinetic analysis of infliximab in patients with ulcerative colitis. *Eur J Clin Pharmacol* 65 (12):1211-1228
38. Kerbel R, Folkman J (2002) Clinical translation of angiogenesis inhibitors. *Nat Rev Cancer* 2 (10):727-739
39. Li J, Wientjes MG, Au JL (2010) Pancreatic cancer: pathobiology, treatment options, and drug delivery. *AAPS J* 12 (2):223-232
40. Kindler HL, Niedzwiecki D, Hollis D, Sutherland S, Schrag D, Hurwitz H, Innocenti F, Mulcahy MF, O'Reilly E, Wozniak TF, Picus J, Bhargava P, Mayer RJ, Schilsky RL, Goldberg RM (2010) Gemcitabine plus bevacizumab compared with gemcitabine plus placebo in patients with advanced pancreatic cancer: phase III trial of the Cancer and Leukemia Group B (CALGB 80303). *J Clin Oncol* 28 (22):3617-3622
41. Rudge JS, Holash J, Hylton D, Russell M, Jiang S, Leidich R, Papadopoulos N, Pyles EA, Torri A, Wiegand SJ, Thurston G, Stahl N, Yancopoulos GD (2007) VEGF Trap complex formation measures production rates of VEGF, providing a biomarker for predicting efficacious angiogenic blockade. *Proc Nat Acad Sci USA* 104 (47):18363-18370

Tables

Table 1. Summary of clinical trials of aflibercept included in the analysis

Study	Aflibercept dose (mg/kg)	Dosing frequency	Concomitant chemotherapy	Sampling scheme frequency	Number of patients with PK data
<i>Phase I</i>					
TED6115/6116: dose-escalation, solid tumors	2, 3, 4, 5, 7	q2wk	single agent	full profile ^a (n=12-15)	37
TCD6118: dose-escalation, combination, solid tumors	2, 4, 5, 6	q2wk	irinotecan/5-FU /LV	full profile ^b (n=19)	61
TCD6120: dose-escalation, combination, solid tumors	2, 4, 5, 6, 7, 9	q3wk	docetaxel	full profile ^c (n=12)	53
<i>Phase II</i>					
ARD6122: dose-escalation, ovarian cancer	2, 4	q2wk	single agent	peak (Cycle 1) & multiple troughs	175
ARD6123: NSCLC	4	q2wk	single agent	peak (Cycle 1) & multiple troughs	77
EFC6125: ovarian cancer with ascites	4	q2wk	single agent	peak (Cycle 1) & multiple troughs	30
<i>Phase III</i>					
VANILLA: combination, pancreatic cancer	4	q2wk	gemcitabine	peak (Cycle 1) & multiple troughs	204
VITAL: combination, NSCLC	6	q3wk	docetaxel	peak (Cycle 1) & multiple troughs	370
VELOUR: combination, colorectal cancer	4	q2wk	FOLFIRI	peak (Cycle 1) & multiple troughs	500

n: number of sampling times per patient, NSCLC: non small cell lung carcinoma, LV: leucovorin, 5-FU: 5-fluorouracil
q2wk/q3wk: every two/three weeks, FOLFIRI: combination of irinotecan.5-FU/LV with different doses to that in TCD6118

^a For study TED6115, samples were taken at pre-dose and at 1, 2, 4, 8, 24, 30, 48, 96, 168 hours post-dose of the first administration; pre-dose, post-dose and 7 days post-dose of the second administration. For study TED6116, an extension of TED6115, samples were taken prior to and at the end of each cycle in all cohorts and then at the end of the study and 3 months after the last dose.

^bThis study was in 2 parts, part 1 open-label, multicenter, dose-escalation design and part 2 started double-blind, multicenter, randomized, parallel group, placebo controlled design. For part 1, samples were taken 5 minutes prior to dosing and at 1, 2, 4, 8, 24, 30, and 48 hours, and 7 days after dosing on Day 1 of Cycles 1 and 2. For all subsequent cycles, samples were collected 5 minutes prior to dosing, and at the end of aflibercept treatment. For part 2, samples were taken 5 minutes prior to dosing for all cycles, and at the end of aflibercept treatment, a final sample was collected.

^cDuring Cycle 1, samples were taken before administration of aflibercept, and at 1, 2, 4, 8, 24 (Day 2), and 48 hours (Day 3) and 7 (Day 8) and 14 (Day 15) days, after the start of aflibercept infusion. For all subsequent cycles, samples were collected only before the administration of aflibercept.

Table 2. Characteristics of the patient study population at the entrance of the trial

	Number of patients (%)	Mean \pm SD	Median (5-95%)
Patient demographics			
Age (years)		59.3 \pm 10.3	60.0 (41.0-75.0)
Weight (kg)		71.8 \pm 16.1	70.0 (49.0-99.1)
< 50 kg	88 (5.8)		
50-100 kg	1344 (89.2)		
\geq 100 kg	74 (4.9)		
Sex			
Male	767 (50.9)		
Female	739 (49.1)		
Ethnicity			
Caucasian	1377 (91.4)		
Black	27 (1.8)		
Asian	75 (5.0)		
Other	27 (1.8)		
Laboratory measurements			
Albumin (ALB)*		0.79 \pm 0.22	0.80 (0.57-0.97)
Alkaline phosphatase (ALK)*		1.18 \pm 0.94	0.86 (0.42- 3.32)
Alanine amino transferase (ALT)*		0.70 \pm 0.53	0.55 (0.22-1.65)
Aspartate amino transferase (AST)*		0.88 \pm 0.62	0.72 (0.33-2.06)
Bilirubin (BIL)*		0.51 \pm 0.26	0.46 (0.20-1.00)
Total protein (TP)*		0.88 \pm 0.10	0.88 (0.74-1.03)
Creatinine clearance (CLCR) (mL/min)		89.9 \pm 31.5	84.2 (47.9-148)
Cancer type			
Colorectal	499 (33.1)		
Non-small cell lung	447 (29.7)		
Pancreatic	204 (13.5)		
Ovarian	205 (13.6)		
Other solid tumors	151 (10.0)		
Cancer treatment			
monotherapy	319 (21.2)		
combined chemotherapy	1187 (78.8)		
docetaxel	423 (28.1)		
irinotecan/5-FU /LV	61 (4.1)		
gemcitabine	204 (13.5)		
FOLFIRI	499 (33.1)		
Dose regimen			
4 mg/kg q2wk	939 (62.3)		
6 mg/kg q3wk	399 (26.5)		
other multiple doses	168 (11.1)		
Pharmacokinetic sampling			
Intensive	151 (10.0)		
Peak and trough	1355 (90.0)		

*The measurements were normalized to the upper normal limit value of each laboratory

Table 3. Parameter estimates of the base and final model

Parameter	Base model		Final model	
	Estimate (RSE %)	IIV (RSE%)	Estimate (RSE %)	IIV (RSE %)
CL_f (L/day)	0.85 (2)	35.2 (3)	0.88 (2)	30.9 (3)
$\beta_{CLf, \text{female}}$	-	-	-0.15 (16)	-
$\beta_{CLf, \text{weight}}$	-	-	0.33 (19)	-
$\beta_{CLf, CLCR}$	-	-	0.18 (20)	-
$\beta_{CLf, ALB}$	-	-	-0.39 (14)	-
$\beta_{CLf, ALK}$	-	-	0.10 (18)	-
$\beta_{CLf, ALT}$	-	-	-0.06 (30)	-
$\beta_{CLf, \text{gemcitabine}}$	-	-	0.09 (37)	-
V_p (L)	3.87 (1)	26.8 (5)	4.35 (2)	22.4 (6)
$\beta_{Vp, \text{female}}$	-	-	-0.21 (12)	-
$\beta_{Vp, \text{weight}}$	-	-	0.39 (16)	-
$\beta_{Vp, CLCR}$	-	-	0.10 (39)	-
Q (L/day)	1.68 (8)	81.9 (9)	1.49 (9)	85.5 (9)
V_t (L)	3.75 (5)	64.2 (4)	3.72 (5)	65.8 (4)
V_b (L)	3.92 (10)	29.1 (5)	4.14 (10)	27.5 (6)
V_{max} (mg/day)	0.77 (10)	-	0.82 (10)	-
K_m ($\mu\text{g/mL}$)	1.79 (7)	41.2 (22)	1.92 (9)	28.2 (42)
$\beta_{Km, \text{female}}$	-	-	-0.23 (35)	-
$\beta_{Km, ALT}$	-	-	0.26 (46)	-
$\beta_{Km, AST}$	-	-	-0.26 (37)	-
$\beta_{Km, \text{gemcitabine}}$	-	-	0.60 (17)	-
CL_b (L/day)	0.18 (10)	23.5 (3)	0.19 (10)	22.3 (3)
$\beta_{CLb, \text{age}}$	-	-	0.18 (26)	-
$\beta_{CLb, CLCR}$	-	-	0.09 (31)	-
$\beta_{CLf, ALB}$	-	-	-0.13 (33)	-
$\beta_{CLf, ALT}$	-	-	-0.08 (27)	-
$\beta_{CLf, \text{irinotecan/5-FU/LV}}$	-	-	-0.13 (26)	-
$\beta_{CLf, \text{docetaxel}}$	-	-	0.06 (30)	-
σ_{fa} ($\mu\text{g/mL}$)	0.04 (6)	-	0.04 (6)	-
σ_{fp} (%)	32.90 (1)	-	32.80 (1)	-
σ_{ba} ($\mu\text{g.eq/mL}$)	0.34 (2)	-	0.34 (2)	-
σ_{bp} (%)	9.01 (4)	-	9.04 (4)	-

Table 4. Assessment of covariate effects on aflibercept pharmacokinetic parameters

Covariate	Quantiles	CL _r (L/day)	% Change†	V _p (L)	% Change†	K _m (µg/mL)	% Change†	CL _b (L/day)	% Change†
Typical patient*		0.88		4.35		1.92		0.19	
Sex	Female	0.76	-14.1	3.53	-18.9	1.53	-20.2	-	
WT	5%: 49	0.79	-11.0	3.78	-13.1	-		-	
	95%: 99.1	1.00	12.0	4.99	14.6				
Age	5%: 41	-		-		-		0.18	-6.6
	95%: 75							0.20	4.1
CLCR	5%: 47.9	0.80	-9.3	4.13	-5.1	-		0.18	-4.6
	95%: 148.1	0.98	11.2	4.60	5.9			0.20	5.2
ALB	5%: 0.57	1.00	14.1	-		-		0.20	4.6
	95%: 0.965	0.82	-7.0					0.19	-2.5
ALK	5%: 0.423	0.82	-7.0	-		-		-	
	95%: 3.233	1.01	14.4						
ALT	5%: 0.222	0.93	5.8	-		1.52	-20.6	0.20	7.6
	95%: 1.646	0.82	-6.6			2.54	32.5	0.17	-8.5
AST	5%: 0.333	-		-		2.35	22.3	-	
	95%: 2.058					1.46	-24.0		
Combination	gemcitabine	0.96	9.2	-		3.49	82.0	-	
	irinotecan/ 5-FU/LV	-		-		-		0.17	-12.5
	docetaxel	-		-		-		0.20	6.5

*: male, 60 years, 70 kg, normalized ALB of 0.80, normalized ALK of 0.86, normalized ALT of 0.55, normalized AST of 0.72, CLCR of 82.2 mL/min, receiving aflibercept as a monotherapy

†: theoretical effect (% change with respect to the typical value) of the covariate considered alone, the other covariate being set to its median value

% changes in PK parameters greater than 20% are presented in bold

Figure legends

Figure 1

Proposed structural model for free and bound aflibercept. Free aflibercept in plasma distributes first to tissues then binds to VEGF to form a complex. Binding to VEGF occurs in the peripheral compartment, follows the law of mass action and can be characterised by a nonlinear equation with MM constants (V_{\max} , K_m). Bound aflibercept (complex) is assumed to be directly eliminated through internalisation (k_{int}).

Figure 2

Goodness-of-fit plots: (top) observed versus predicted concentrations for free aflibercept; (bottom) observed versus predicted concentrations for bound aflibercept. The plots on the left present observed versus the population predicted concentration (PRED) while the plots on the right present observed versus individual predicted concentrations (IPRED). One outlier was removed from the plots.

Figure 3

Normalised prediction distribution error (NPDE) vs time since last dose (TimeL) for free aflibercept (top) and bound aflibercept (bottom) with 95% prediction intervals around the 10th, 50th, 90th percentiles. NPDE plots for TimeL \leq 22 weeks are presented in the left and for TimeL \leq 6 weeks in the right. Observed data are plotted using a circle (\circ). The solid line represents the 10th, 50th and 90th percentiles of the npde corresponding to observed data. The shaded area represents 95% prediction intervals for the selected percentiles (pink for 50th percentiles and blue for others).

Figure 4

Examples of individual fits of free and bound aflibercept for 4 subjects in different studies. From left to right: study TED6115 (phase I), study ARD6123 (phase II), study Vital (phase III), and study VELOUR (phase III). Fits for free aflibercept are presented in the top, bound aflibercept in the bottom. Observed data are plotted using a circle (\circ). The line (—) represents the prediction of model.

Figure 5

Predicted median profiles of bound aflibercept at steady-state with different dosing regimens (2, 4, 6 mg/kg q2wk or q3wk) illustrated by vertical lines to x-axis (blue for q2wk and grey for q3wk dosing regimen).

Figure 6

Predicted steady-state concentrations of free aflibercept (left y-axis) and bound aflibercept (right y-axis) following 8 aflibercept doses of 4mg/kg q2wk during 24 weeks.

Figures

Figure 1.

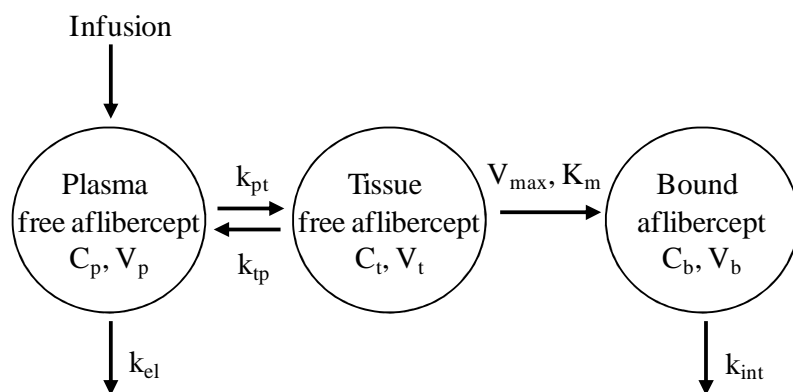


Figure 2.

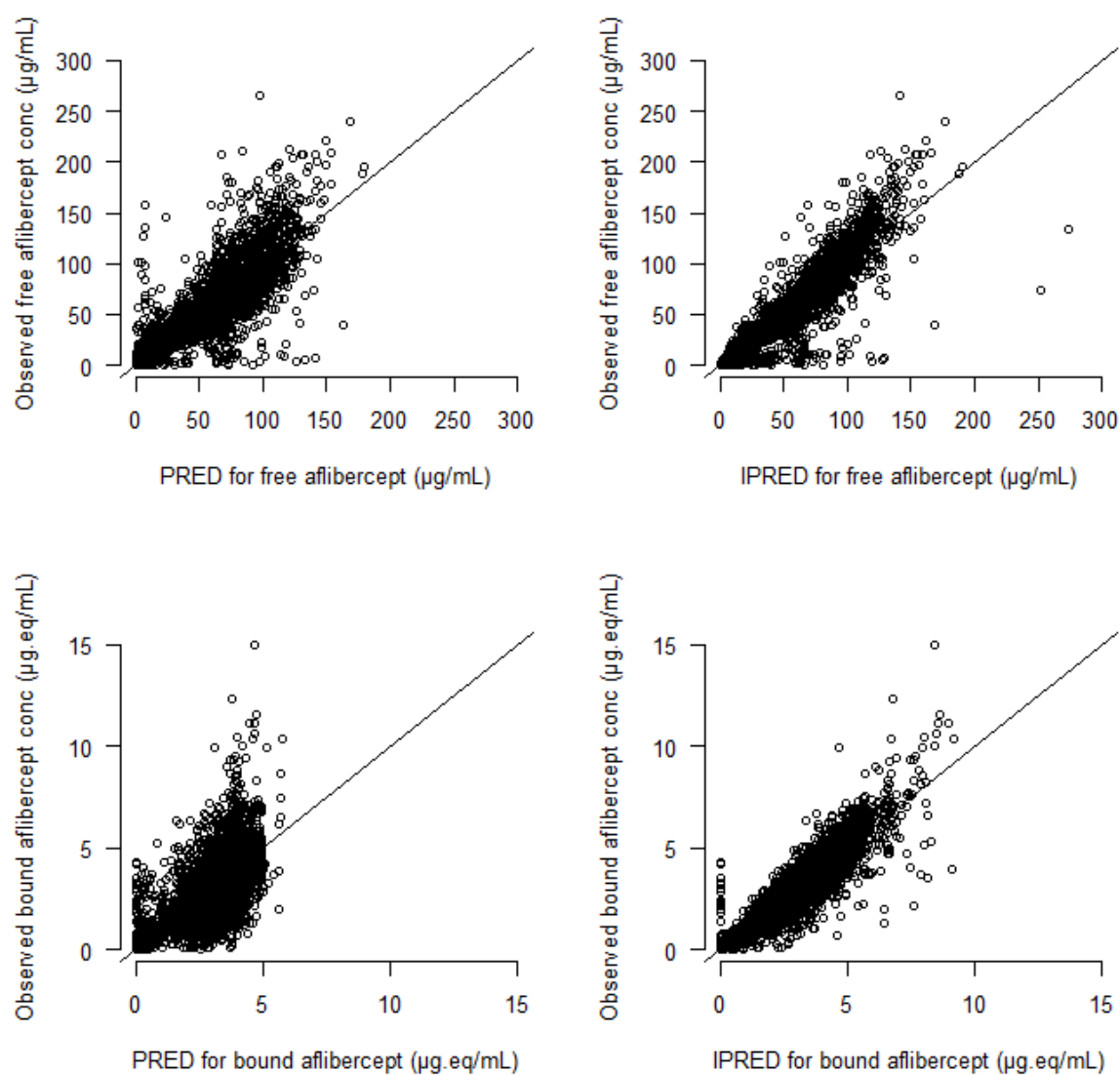


Figure 3.

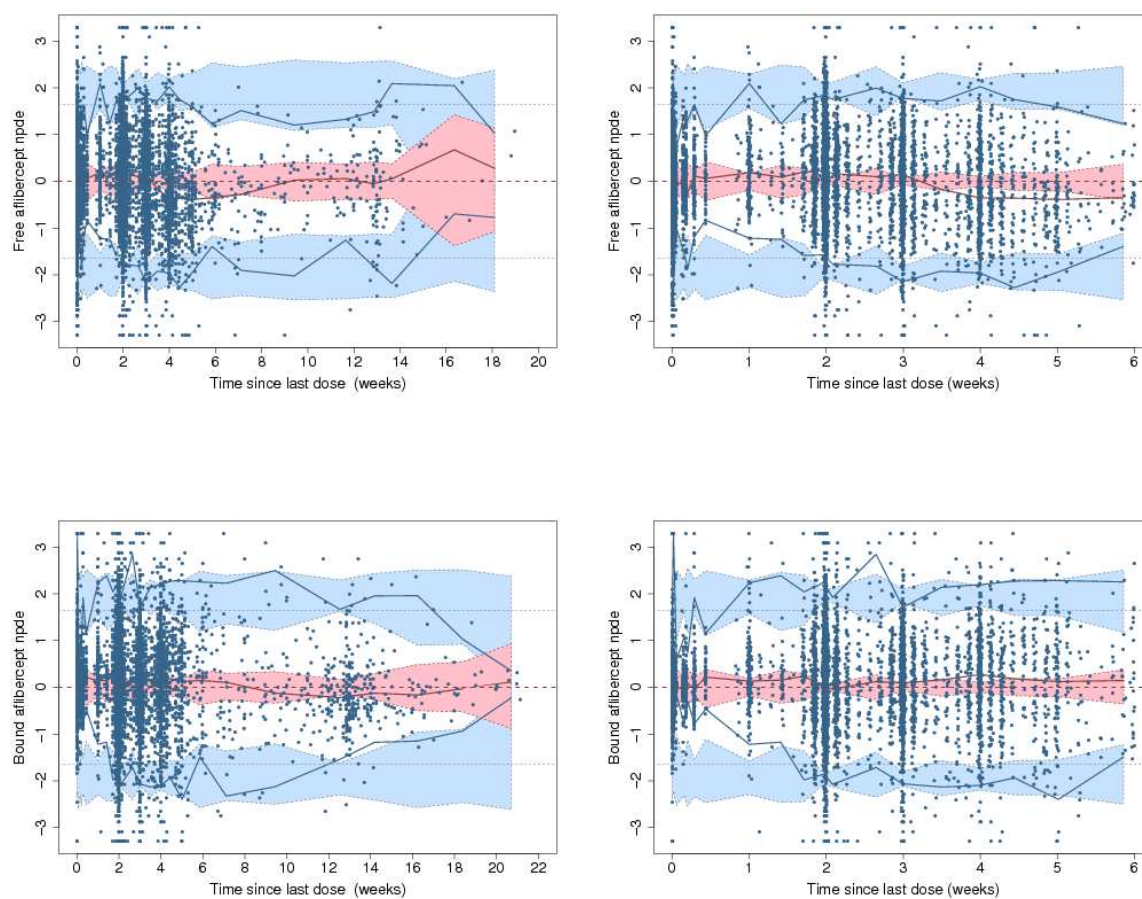


Figure 4.

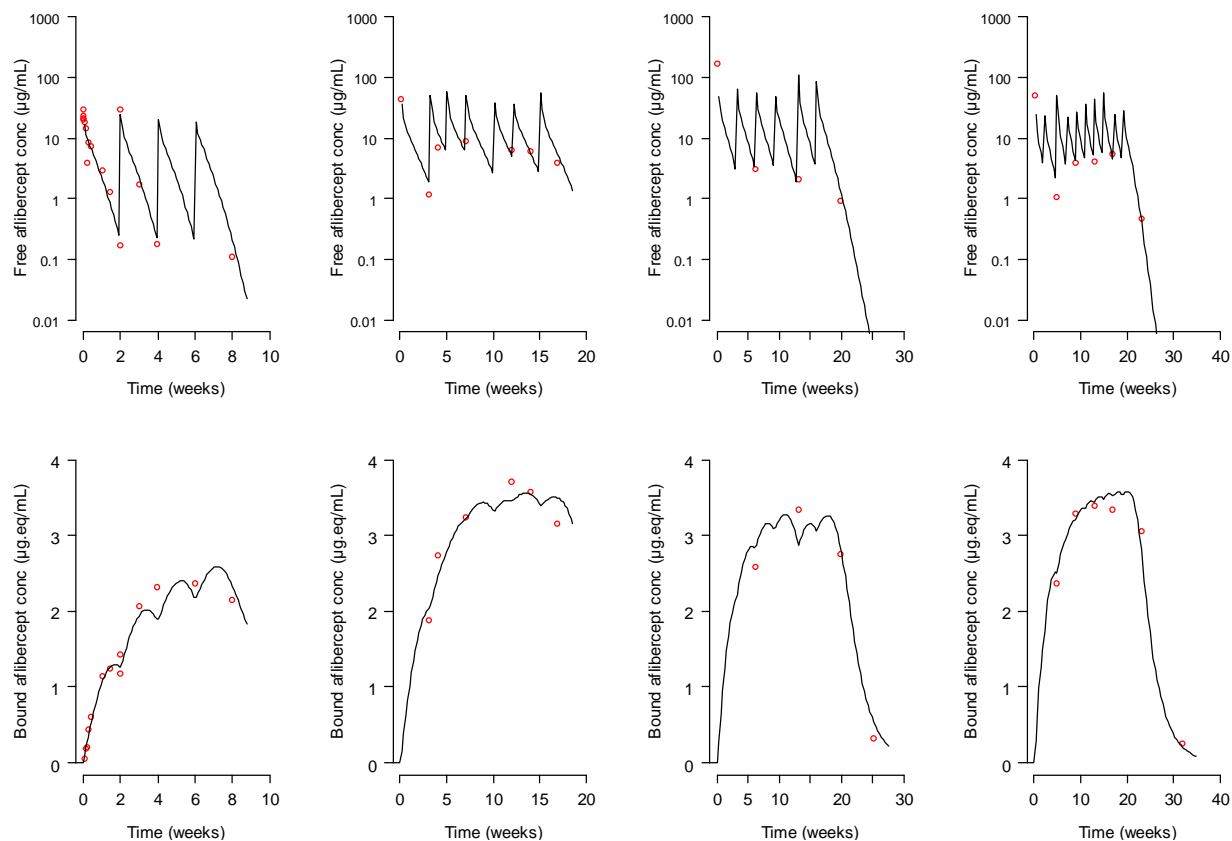


Figure 5.

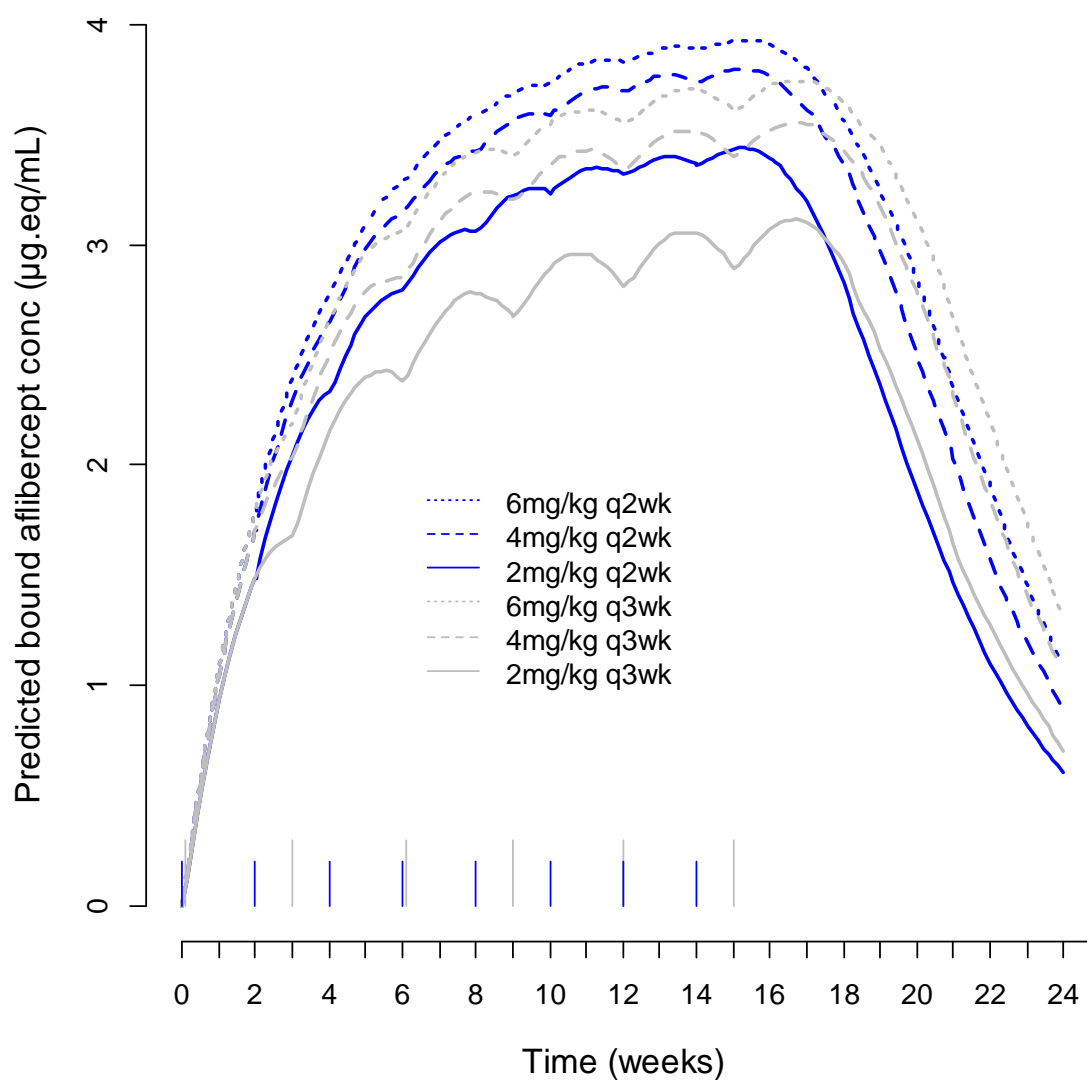
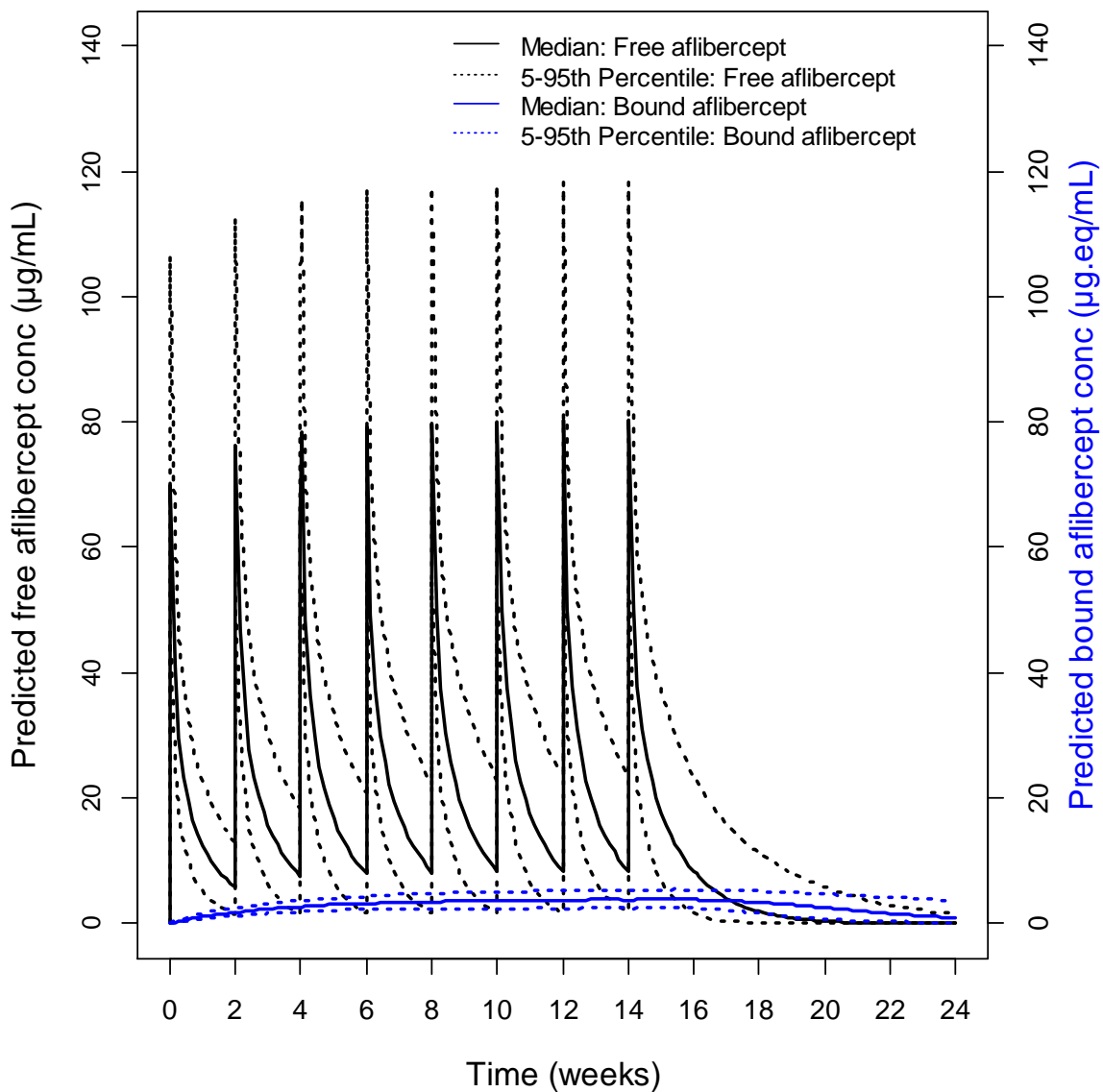


Figure 6.



2.3 Modélisation des données longitudinales de réponse tumorale et de l'exposition à l'affibercept chez les patients atteints de cancer du côlon

2.3.1 Résumé

La modélisation PK dans des travaux précédents nous a permis de caractériser le mécanisme sous-jacent de l'affibercept sur le VEGF chez les patients atteints de cancer. Grâce aux concentrations de l'affibercept lié, elle a également permis de déterminer les schémas d'administration qui commencent à saturer le VEGF. Cependant, les concentrations de l'affibercept lié mesurées dans plasma ne sont qu'un faible biomarqueur PD de l'efficacité du produit. Selon son mécanisme d'action, l'affibercept agit en diminuant la vascularisation de la tumeur et donc sa capacité de grossir. De ce fait, la croissance de la tumeur est pour ce produit une mesure PD plus appropriée pour rendre compte de son activité. Nous avons ensuite, dans ce travail, développé un modèle PD caractérisant l'effet de l'affibercept sur la croissance tumorale pour mieux évaluer son efficacité.

Pour ce faire, nous avons utilisé les données longitudinales de la taille de la tumeur issues de l'étude VELOUR. C'était une étude de phase III multicentrique, randomisée, contrôlée par placebo dans le but de comparer l'efficacité de l'affibercept par rapport au placebo en association avec la chimiothérapie FOLFIRI (5-FU, la leucovorine et l'irinotécan) chez des patients atteints de cancer colorectal métastatique préalablement traités avec l'oxaliplatine. Les patients ont été randomisés pour recevoir soit l'affibercept (4 mg/kg par voie intraveineuse) soit un placebo toutes les 2 semaines en association avec FOLFIRI. Un total de 1069 patients, ayant eu au moins une évaluation de la tumeur avant (baseline) et après le début du traitement, ont été inclus dans l'analyse avec 529 patients dans le groupe de l'affibercept (l'affibercept + FOLFIRI) et 540 patients dans le groupe de référence (placebo + FOLFIRI). La construction du modèle a été basée sur le modèle d'inhibition de la croissance tumorale (TGI) développé par Claret et al pour le bévacizumab chez les patients atteints du cancer colorectal (CLARET et al., 2009). Ce modèle prend en compte la dynamique tumorale, l'effet du médicament antitumoral et la résistance à l'effet du médicament. Pour tenir compte du mécanisme d'action de l'affibercept et du FOLFIRI, nous avons supposé que FOLFIRI stimule la dégradation des cellules tumorales (effet sur le paramètre K_D représentant la vitesse de disparition des cellules tumorales) tandis que l'affibercept lié (le complexe affibercept-VEGF) inhibe la croissance des cellules tumorales (effet sur le paramètres K_L représentant la vitesse de croissance tumorale). L'affibercept libre n'est pas pris en compte dans le modèle d'effet mais il sert à générer la PK du lié. Une approche de modélisation PK/PD séquentielle a été utilisée pour les données pharmacocinétiques de l'affibercept. Les paramètres

individuels des patients dans les données VELOUR ont été calculés à partir du modèle TMDD développé dans l'analyse pharmacocinétique de population précédente avec les données issues de 9 essais cliniques. Ils ont ensuite été utilisés pour simuler les profils de concentrations de l'affibercept lié. La prise en compte de l'effet FOLFIRI a été introduite différemment dans le modèle en l'absence de données PK de la combinaison 5-fluorouracile, la leucovorine et l'irinotécan. Le FOLFIRI a été considéré comme un médicament virtuel et son historique de doses a été introduit dans le modèle TGI en utilisant une approche de modélisation K-PD (*kinetic-pharmacodynamic*). Un modèle monocompartimental iv bolus virtuel a été supposé pour la PK du FOLFIRI.

Ce modèle TGI a décrit correctement l'évolution de la taille de la tumeur dans le groupe de l'affibercept et dans le groupe de référence (placebo). Cependant, les données de sortie d'étude n'ont été prises en compte ni modélisées dans ce modèle. Cela pourrait biaiser les estimations des paramètres. La prochaine étape consisterait à les prendre en compte par modélisation conjointe des données longitudinales et des données de survie.

Ce travail est présenté dans un rapport technique.

2.3.2 Rapport technique

Longitudinal PK/PD modeling of tumor response and exposure to aflibercept in colorectal cancer

Introduction

Evaluation of treatment efficacy using the longitudinal tumor size models has become important for the development of antitumor drugs [1-3]. Several models of tumor growth inhibition (TGI) have been developed, from the simple empirical model with mixed exponential-decay (shrinkage) and linear-growth (progression) components [2] to complex tumor dynamics models incorporating drug and disease-specific parameters [4-6]. In colorectal cancer, a semi-mechanistic exposure-driven TGI model was able to use the data of tumor dynamics on a phase II trial of capecitabine to predict overall survival in a subsequent phase III trial [5]. Recently, a simplified version of this TGI model assuming constant exposure over time was used to capture treatment effect and predict survival from two phase III studies comparing bevacizumab plus chemotherapy versus chemotherapy [7]. Neither the PK data nor dosing intensity history of the treatment was modeled in the simplified model.

Aflibercept (Zaltrap®) is a fusion protein consisting of human vascular endothelial growth factor (VEGF) receptor extracellular domains (domain 2 from VEGFR1 and domain 3 from VEGFR2) fused to the Fc portion of human Immunoglobulin G1 (IgG1). It interferes with the biological actions of VEGF by forming a complex with VEGF in the blood stream and extravascular space and preventing it from interacting with its receptors on endothelial cells. As an alternative strategy for inhibiting tumor growth, this novel antiangiogenic agent is being developed worldwide for the potential treatment of solid tumors [8,9]. The VELOUR Phase III trial evaluating aflibercept as a second-line treatment for metastatic colorectal cancer (aflibercept in combination with FOLFIRI (leucovorin, 5-fluorouracil, and irinotecan) versus FOLFIRI plus placebo has demonstrated the benefits of aflibercept on progression-free survival and overall survival [10]. This indication has been recently approved in the United States and Europe.

In previous studies, the pharmacokinetics (PK) of aflibercept and its binding to VEGF were well characterized by a target-mediated drug disposition (TMDD) model with irreversible Michaelis-Menten binding in healthy subjects [11] and patients with advanced solid tumor [12]. This model allowed predicting the individual concentrations of free and bound

aflibercept of the patients in the VELOUR trial. However, the pharmacodynamic (PD) aspects of aflibercept in colorectal cancer, especially the relation between the PK, the binding to VEGF and the effect on tumor growth, which may help to predict the efficacy of aflibercept have not been studied yet. The objective of this analysis was to develop a longitudinal PK/PD model of tumor response and exposure to aflibercept and FOLFIRI with data from the VELOUR trial.

Methods

Study design

As described in [10], the VELOUR trial was a multicenter, randomized, placebo-controlled Phase III trial with the goal of comparing the efficacy of aflibercept versus placebo in combination with the FOLFIRI regimen as second-line treatment for patients with metastatic colorectal cancer (MCRC) previously treated with oxaliplatin. Patients were randomized to receive either aflibercept (4 mg/kg intravenously, 612 patients) or placebo (614 patients) every 2 weeks in combination with FOLFIRI. The primary endpoint was overall survival (OS). Secondary endpoints included progression free survival (PFS), overall response rate (RR), toxicity, immunogenicity and PK of aflibercept. Dose adjustments for each treatment group and/or cycle delays were permitted in the event of toxicity.

A total of 1069 patients with at least one tumor evaluation after the baseline were included in the analysis with 529 patients in the aflibercept group (aflibercept+FOLFIRI) and 540 patients in the reference group (placebo+FOLFIRI).

PK /PD sample collection

Plasma samples for the assessment of aflibercept PK were collected pre-dose, at the end of the infusion of the first cycle, then, prior to administration of study treatment in each odd numbered cycle and approximately 30 and 90 days after the last infusion of aflibercept/placebo. Free and VEGF-bound aflibercept concentrations were measured by a validated enzyme linked immunosorbent assay (ELISA) method.

For the assessment of tumor response, tumor imaging using magnetic resonance imaging technique (MRI) was performed prior to each treatment administration (baseline), then every

6 weeks during the treatment period and approximately 30 days after the last infusion of aflibercept/placebo. The sums of longest diameter of the target lesions were used as the continuous drug-response data for PD modeling.

Population PK model for aflibercept

A TMDD model with irreversible binding Michaelis-Menten approximation was used to describe the plasma concentrations of free and VEGF-bound aflibercept versus time data collected from 9 Phase I-III clinical trials including the VELOUR trial [12]. With very limited PK sample collection, the data of the VELOUR trial was too sparse to be modeled alone. This structural model is shown in Figure 1.

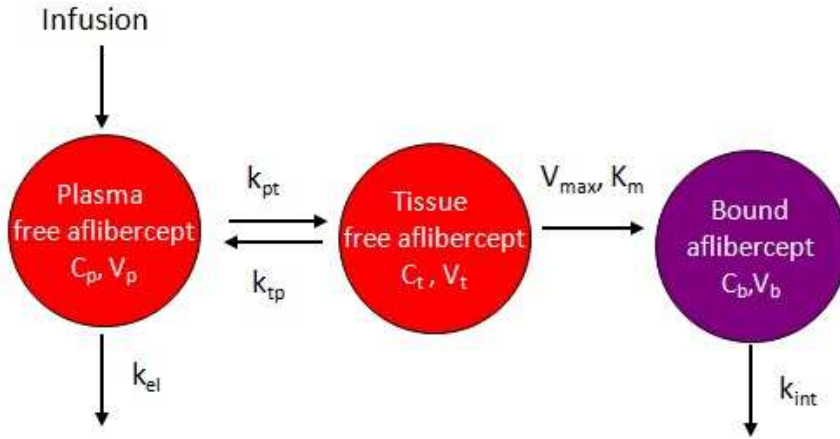


Figure 1. PK model for free and VEGF-bound aflibercept

In this model, the concentration of free aflibercept in central compartment (C_p), in tissue compartment (C_t) and the concentration of bound aflibercept (C_b) are described by the following system of differential equations:

$$\begin{aligned} \frac{dC_p}{dt} &= \frac{R_{inf}}{V_p} - (k_{el} + k_{pt})C_p + k_{tp} \frac{C_t V_t}{V_p} \\ \frac{dC_t}{dt} &= k_{pt} \frac{C_p V_p}{V_t} - k_{tp} C_t - \frac{1}{V_t} \cdot \frac{V_{max} C_t}{K_m + C_t} \\ \frac{dC_b}{dt} &= \frac{1}{V_b} \cdot \frac{V_{max} C_t}{K_m + C_t} - k_{int} C_b \end{aligned}$$

where R_{inf} is the infusion rate, k_{el} (day^{-1}) is the first order elimination rate constant of free aflibercept from the central compartment, k_{tp} and k_{pt} (day^{-1}) are the first order rate constants between the central and the peripheral compartment, k_{int} (day^{-1}) is the first order rate constant of bound aflibercept internalization, V_p (L) is the central volume of distribution of free aflibercept, V_t (L) is the peripheral volume of distribution of free aflibercept, V_b (L) is the volume of distribution of bound aflibercept, V_{max} (mg/day) is the maximum binding capacity, K_m ($\mu\text{g/mL}$) is the concentration of free aflibercept corresponding to half of the maximum binding capacity.

The clearances of free and bound aflibercept (CL_f and CL_b) and the intercompartmental clearance of free aflibercept (Q) between the central and the peripheral compartments are derived from the micro-constant as follows:

$$\begin{aligned} CL_f &= k_{el} \cdot V_p \\ CL_b &= k_{int} \cdot V_b \\ Q &= k_{tp} \cdot V_t = k_{pt} \cdot V_p \end{aligned}$$

The parameters to be estimated in this model are: CL_f , V_p , Q , V_t , CL_b , V_b , V_{max} , and K_m .

With respect to the law of mass action, the concentrations of bound aflibercept were converted into equivalent concentrations of free aflibercept by multiplying them with 0.717, the ratio of molecular weights between free and bound aflibercept. The units of free aflibercept and bound aflibercept concentrations were $\mu\text{g/mL}$ and $\mu\text{g.eq/mL}$ respectively.

An exponential interindividual model implying a log-normal distribution was included on all parameters except V_{max} . The variance-covariance matrix was modeled using a diagonal matrix. The residual variability was modeled using a combined additive and proportional model for both free and VEGF-bound aflibercept.

The covariate analysis was performed after obtaining the base model and focused on the main parameters describing the PK of free and VEGF-bound aflibercept and their binding to VEGF: CL_f , V_p , V_{max} , K_m and CL_b . We found that CL_f and V_p increased with body weight. Patients with low serum albumin (ALB) or high serum alkaline phosphatase (ALK) had approximately 14% faster CL_f compared to a typical patient. Compared to men, women had 14.1 % slower CL_f , 18.9 % lower V_p and 20.2 % lower K_m . Pancreatic cancer patients

(confounded with effect of gemcitabine) had 82% larger K_m . The contribution of these covariates on explaining the variability of PK parameters was however small.

The individual PK parameters for patients in the VELOUR trial were obtained as post hoc estimates from the final PK model with covariates for these 9 clinical phase I-III trials. For the evaluable patients without PK data, their PK parameters were calculated using the typical PK parameters and their own covariates.

Kinetic-pharmacodynamic (K-PD) model for FOLFIRI

Due to lack of PK data, the dosing history of FOLFIRI was modeled using a kinetic-pharmacodynamic (K-PD) approach [13]. Three drugs in the chemotherapy (leucovorin, 5-fluorouracil, and irinotecan) were not modeled separately. FOLFIRI was considered as a whole drug with the virtual dose calculated as the sum of doses of three drugs. The PK of FOLFIRI is then described by a simple virtual one-compartment model (aimed to represent the biophase) with bolus input. Parameter for this model is the elimination rate constant from the virtual compartment (k_{pF}), which describes the equilibrium between the rate of dose administration and the observed effect.

Population PK/PD model

A tumor growth inhibition (TGI) model accounting for the dynamics of tumor growth, antitumor drug effect and resistance to drug effect was developed by Claret et al [5]. This model is described by the differential equation below:

$$\frac{dTS(t)}{dt} = K_L \cdot TS - K_D \cdot Exposure(t) \cdot e^{-\lambda t} \cdot TS \quad TS(0) = TS_0$$

where TS is the tumor size at time t , TS_0 is the baseline tumor size, K_L is the tumor growth rate, K_D is the drug constant-cell-kill rate, λ is the rate constant of resistance appearance and exposure (t) is the drug exposure at time t .

A mechanism-based PK/PD model inspired from the above TGI model was proposed to describe the underlying disease progression and exposure-driven drug effect of both

aflibercept and FOLFIRI on tumor size over time. According to the mechanism of action, aflibercept effect was identified by its inhibition on the tumor growth rate K_L whereas FOLFIRI effect was identified by its stimulation of the tumor cell kill rate K_D . The structural model for this TGI model is shown in Figure 2.

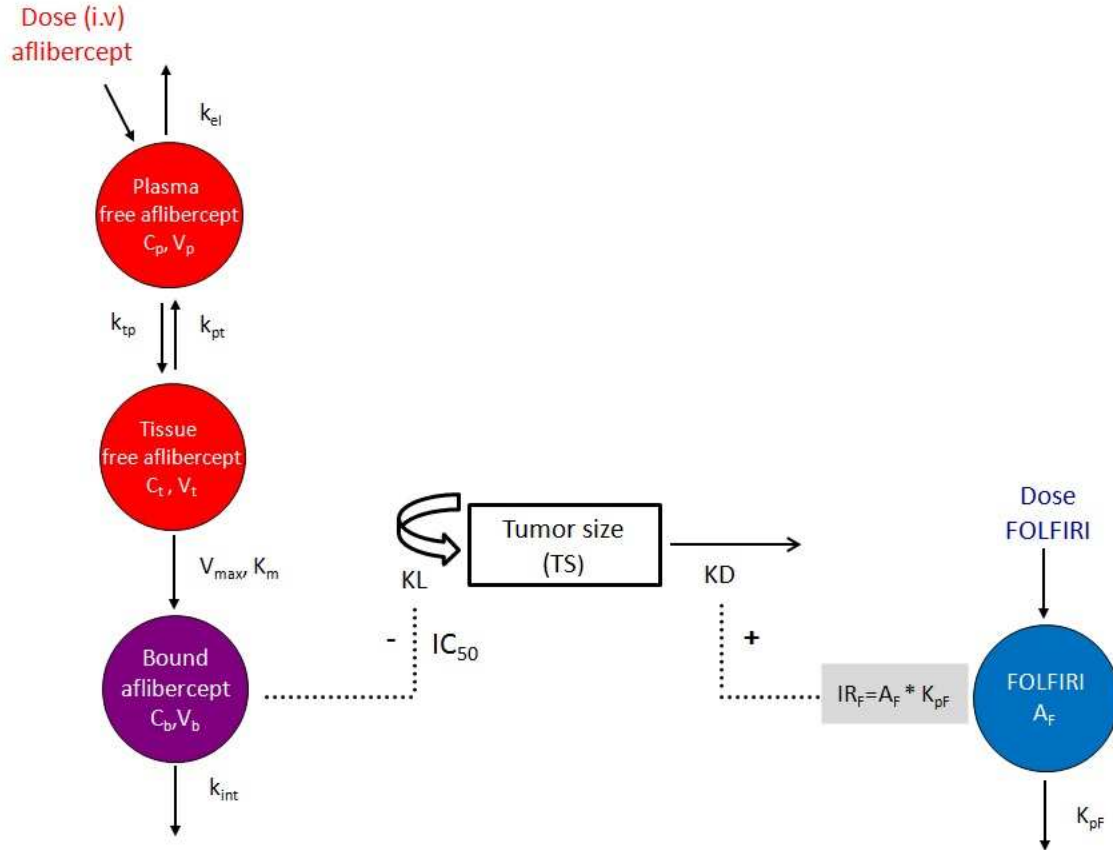


Figure 2. The structure PK/PD model for tumor size

In this model, the exposure of aflibercept which inhibits the tumor growth rate is given by the VEGF-bound aflibercept concentrations (C_b) which were calculated using the individual PK parameters (sequential PK/PD modeling approach). The exposure of FOLFIRI which stimulates the tumor cell-kill rate is given by a virtual infusion rate (IR_F) from dosing compartment of FOLFIRI, expressed in drug amount of FOLFIRI (A_f) per time unit.

The proposed TGI model is described by the differential equations as below:

$$\frac{dA_f}{dt} = -K_{pF} \cdot A_f$$

$$IR_F = k_{pF} \cdot A_f$$

$$\frac{dTS}{dt} = K_L \cdot \left(1 - \frac{C_b}{C_b + IC_{50}}\right) \cdot TS - K_D \cdot IR_F \cdot e^{-\lambda t} \cdot TS$$

$$TS(0) = TS_0$$

where IC_{50} is the concentration of bound aflibercept corresponding to half of effect of aflibercept in controlling the tumor growth and k_{pF} is the elimination rate from the virtual compartment.

The use of free aflibercept concentrations instead of bound aflibercept concentrations was also investigated in this study.

Due to the difficulty to identify all the parameters of the model, we initially chose to fix the k_{pF} parameter to 1.73 week^{-1} . This value allows the effect of FOLFIRI not to be accumulated after a 2 week dosing interval which is equal to 5 times of the virtual half-life.

An exponential interindividual model implying a log-normal distribution was included on all parameters. The variance-covariance matrix was modeled using a diagonal matrix. The residual variability was modeled using an additive model.

Population PK/PD data analysis

The population PK/PD analysis was performed using a nonlinear mixed-effects modeling approach with MONOLIX 4.1.2 implementing the SAEM algorithm. The model control files were written using MLXTRAN script. Graphical analyses for model evaluation were performed with R 2.14.1.

The likelihood ratio test was used to discriminate between nested models through the difference in log likelihood (-2LL). A p-value of 0.05 was considered statistically significant. For non-nested models, the model selection was based on the Bayesian information criterion BIC. The better model is the one with a smaller value of BIC

Model evaluation

Internal evaluation of the model was based on goodness-of-fit (GOF) plots, including plots of observations versus individual and population predictions, plots of NPDE [14,15] versus time and predictions. Plots of visual predictive check (VPC) were used to assess model predictive performance, based on the simulation of 1000 datasets.

Results

Patient

The database for this population PK/PD analysis consisted of 1069 patients involved in the VELOUR trial. These patients had at least one tumor evaluation after the start of study treatment. There were 529 patients in the reference group and 540 patients in the aflibercept group, of whom 445 patients had PK data. Approximately 5 observations were available for each patient (range: 2 to 17 observations). A total of 5072 tumor size measurements were used for model building in this analysis. The characteristics of the patients in the two groups have been previously published [10].

Figure 3 presents the observed tumor size versus time of 16 subjects taken at random equally in the reference group and the aflibercept group.

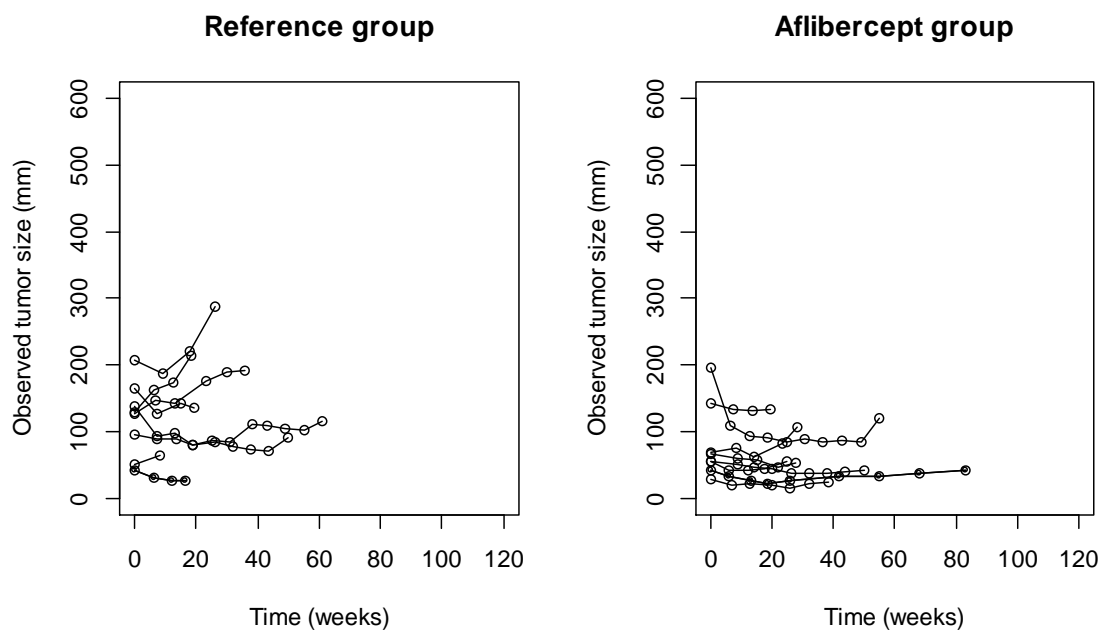


Figure 3. Observed tumor size versus time in aflibercept and reference group

The time course of tumor size in both groups shows various profiles with a large variability. Each patient had average 4-5 measurements. There was also a large heterogeneity in the observed data on the duration of tumor measurements. The distribution of the last time at which tumor was measured for all subjects is shown in Figure 4.

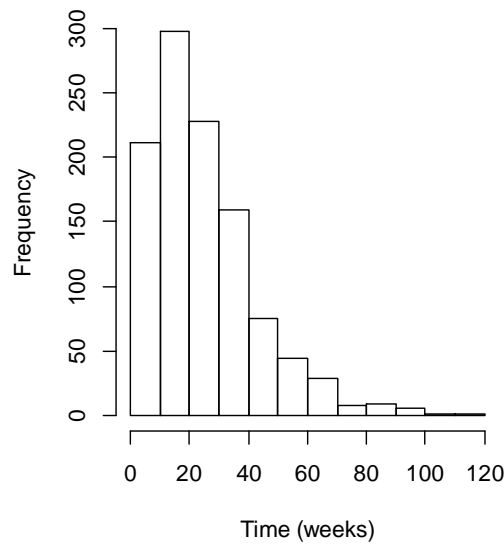


Figure 4. Histogram of last time of tumor measurement

About 47.7% patients having the last tumor size measurement at time ≤ 20 weeks, 36.2% at time between 20 and 40 weeks and 16.1% at time > 40 weeks. The difference in last time of tumor measurement was due to the drop-out during the study period.

The estimated parameter values for the proposed TGI model are presented in Table 1.

Table 1. Parameter estimates for the TGI model

Parameter	Estimate	RSE
TS_0 (mm)	91.4	2
K_L ($\mu\text{g.eq/mL.week}^{-1}$)	0.0123	5
IC_{50} ($\mu\text{g.eq/mL}$)	1.77	48
K_D (week^{-1})	1.32e-05	4
λ (week^{-1})	0.0569	7
k_{pF} (week^{-1})	1.73 (fixed)	-
ω_{TS_0} (%)	74.7	2
ωK_L (%)	83.5	4
ωIC_{50} (%)	180	32
ωK_D (%)	70.6	4
$\omega \lambda$ (%)	85.3	7
ωk_{pF} (%)	-	-
σ_a (mm)	7.58	1

All model parameters were well estimated with relative standard errors (RSE) reported to be less than 10%, except for IC_{50} and $\omega_{IC_{50}}$ with RSE of 30-50%. The typical concentration of VEGF-bound aflibercept corresponding to half inhibition effect was estimated to be 1.77 ($\mu\text{g.eq/mL}$), a half of observed plateau concentrations for VEGF-bound aflibercept. The interindividual variability of model parameters was high, ranging from 74.7% (TS_0) to 180% (IC_{50}). The additive error was small compared to the observed tumor sizes. The η -shrinkage was large for all parameters except for TS_0 (7.1%), reflecting the lack of information in the patients: K_L (51.2%), IC_{50} (94.8%), K_D (55.4%), λ (83.4%). The ε -shrinkage was also large (50.4 %).

The goodness-of-fit plots of the TGI model are shown in Figure 5. The plots of observations versus population predictions and observations versus individual predictions indicated that the model adequately described the observations. The plots of NPDE however show a tendency for late times and high tumor sizes. This bias was due to a large amount of drop-out observations during the study period.

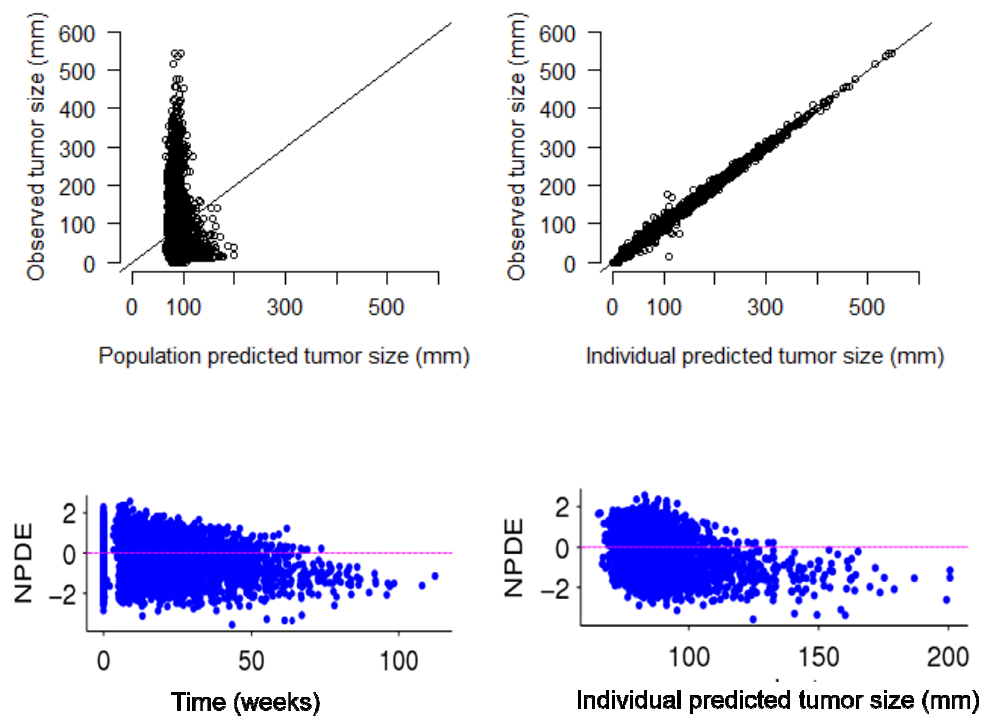


Figure 5. Goodness-of-fit of the TGI model with observed tumor size vs population or individual predicted tumor size plots (on the top) and normalized prediction distribution error (NPDE) versus time and individual predicted tumor size (on the bottom)

Examples of individual fits taken at random for the reference and the aflibercept group are shown in Figure 6. This complex TGI model adequately described the data and the model is flexible enough to describe various types of individual profiles and accompany both reductions and subsequent regrowth of tumor size.

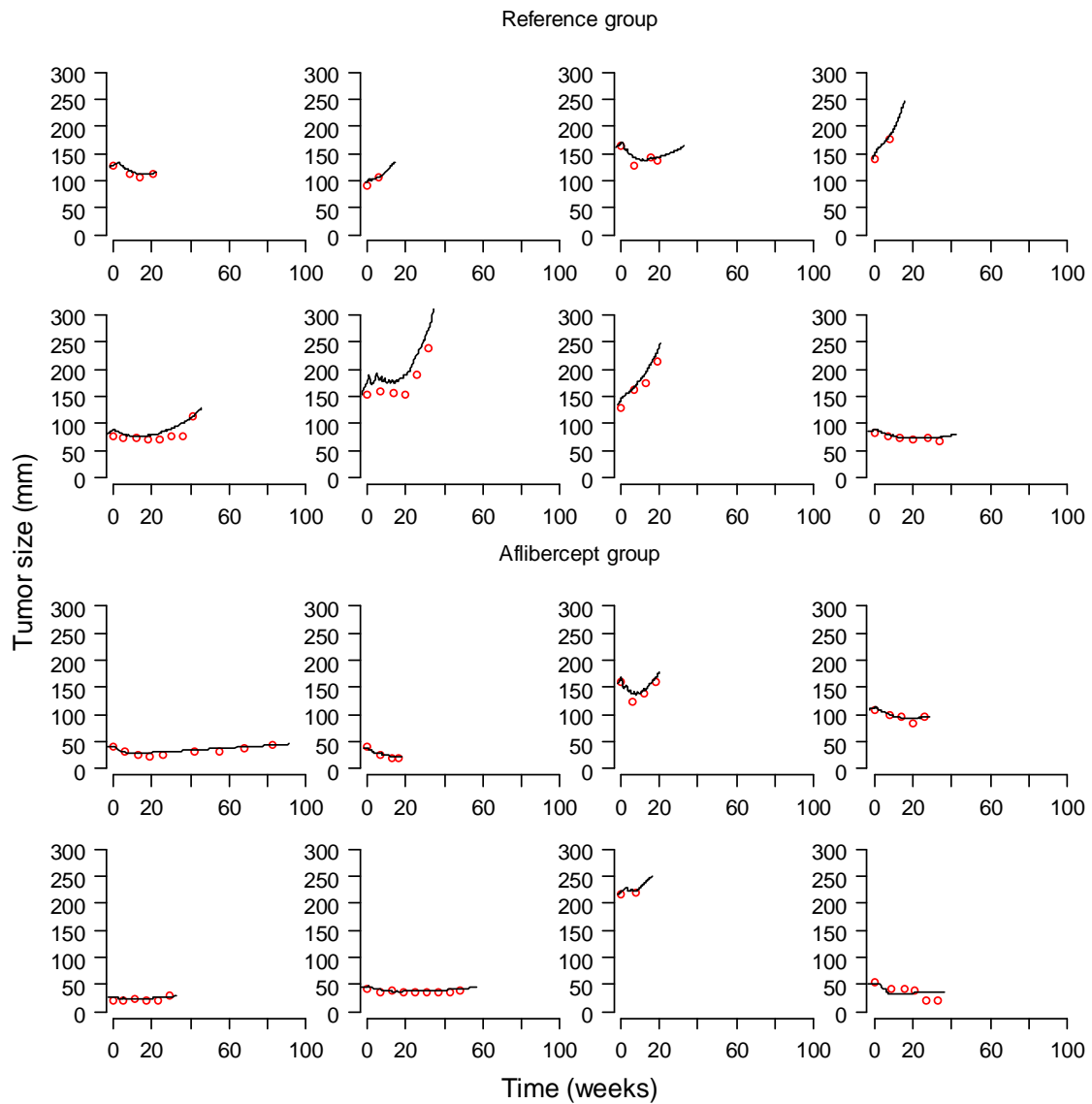


Figure 6. Example of individual fits of the TGI model.

The visual predictive check (VPC) plots of the proposed TGI model are shown in Figure 7 for the whole study period as well as for different period after the first tumor evaluation: baseline to 15 weeks, 15 to 30 weeks, and 30 to 45 weeks.

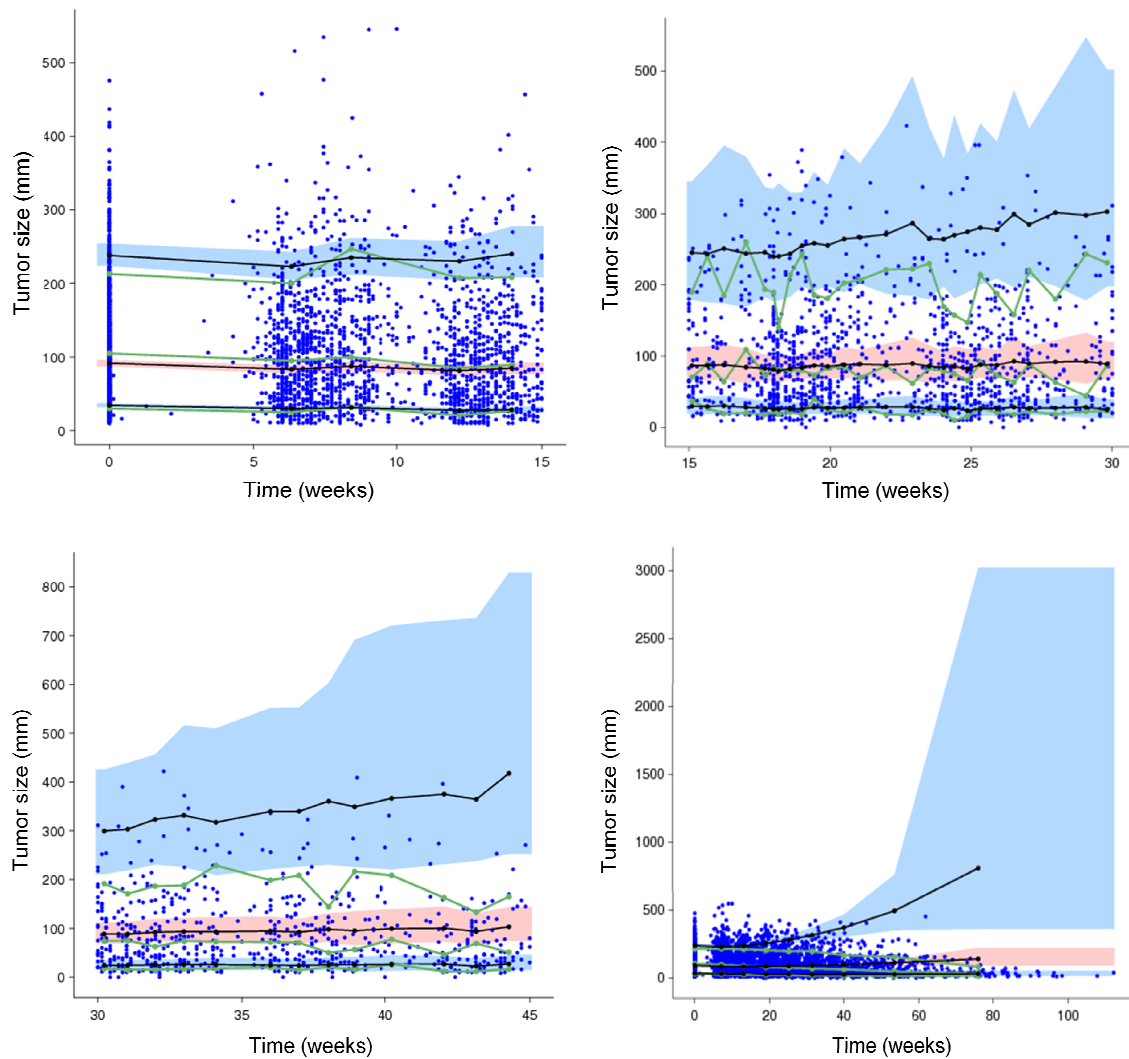


Figure 7. Visual predictive check for the TGI model. Observed data are plotted using a dot (\bullet). The shaded area and the black solid lines represent the 90% prediction interval and the predicted median of 10th, 50th and 90th percentiles of simulated data (n=1000). The green solid lines represent the 10th, 50th and 90th percentiles of observed data.

The model predicts well the tumor sizes at early times, especially from the baseline to 30 weeks. However, the predicted tumor sizes from 60 weeks to the end of the study were extremely large, compared to the observations. This disagreement between observations and predictions from 1000 simulated data may due to the large variability obtained for all model parameters and drop-out during the late times of the study.

Figure 8 displays the predicted profiles of tumor size for a typical patient in VELOUR study in the aflibercept group and the reference group receiving the treatments during 6 months. The

tumor is shown to be better controlled in the aflibercept group with lower sizes during the treatment time and a later regrowth phrase compared to the reference group.

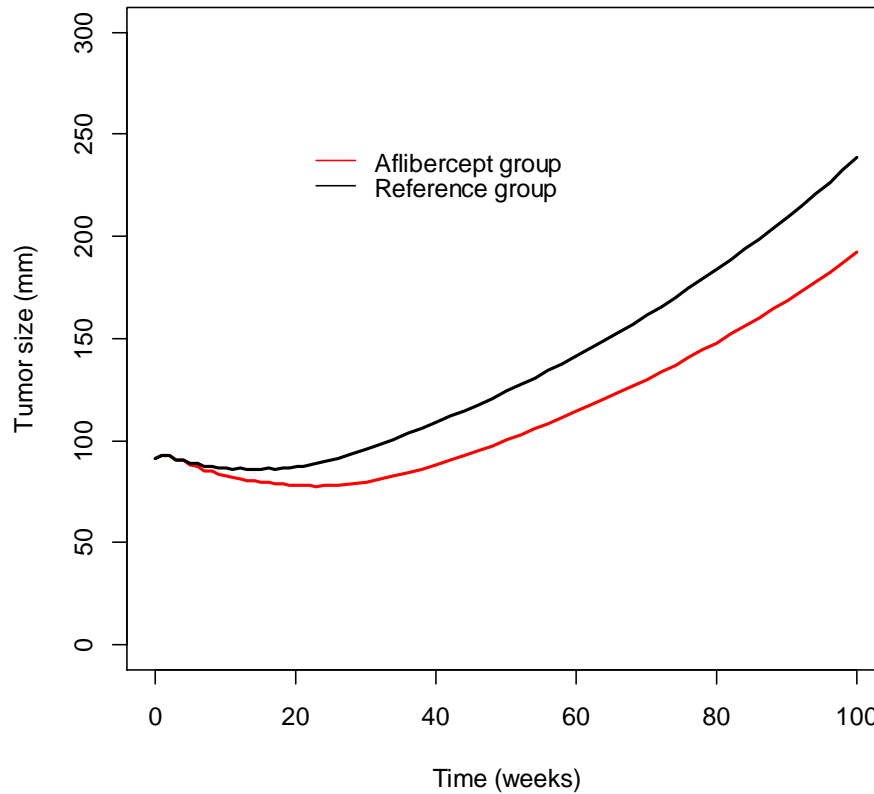


Figure 8. Predicted profile of tumor size for a typical patient in VELOUR study receiving 5000 mg FOLFIRI (virtual dose) and 4 mg/kg aflibercept every 2 weeks during 6 months (aflibercept group) or receiving only 5000 mg FOLFIRI every 2 weeks during 6 months (reference group)

Discussion

In this study, we report the development of a population PK/PD model describing the time-course of tumor size in relationship with the exposure-driven effects of aflibercept given intravenously 4 mg/kg every 2 weeks in combination with FOLFIRI in 1069 evaluable patients with MCRC from the VELOUR trial. The free and bound aflibercept concentrations were measured as well as the tumor sizes.

The structural TGI model describing the time-course of tumor size was inspired from the model of Claret et al [5] where the tumor dynamics is controlled by the tumor growth rate K_L

(disease-specific parameter), the drug-constant cell kill rate K_D and the resistance constant rate λ (drug-specific parameters). To introduce the exposure of Aflibercept and FOLFIRI in this TGI model, we used two approaches. A sequential PK/PD modeling approach was used for PK data of aflibercept. The individual parameters of the patients in the VELOUR data was calculated from the TMDD model developed in a previous population PK analysis with data pooled from 9 clinical trials. They were then used to simulate the profiles of VEGF-bound aflibercept concentrations which increase from the treatment start until reaching the plateau, resulting in a reduction in tumor size and then decrease after the end of last treatment dose. The introduction of FOLFIRI data in the TGI model is however different to that of aflibercept because there was no PK data collected. It is even more complex since FOLFIRI is the combination of three drugs (leucovorin, 5-fluorouracil, and irinotecan). The dose history of FOLFIRI considered as a virtual drug was indeed introduced in TGI model using a K-PD modeling approach assuming one compartment virtual i.v bolus model for PK of FOLFIRI.

Extending the TGI model for an antiangiogenic drug, the exposure of aflibercept was introduced as inhibition of K_L instead of simulation of K_D . This is consistent with its mechanism of action which binds to VEGF and inhibits the tumor growth. Contrarily, the exposure of FOLFIRI was incorporated as stimulation of K_D since the chemotherapy has more direct effect on tumor size by killing tumor cells.

The choice between free and bound aflibercept concentrations to be involved in the TGI model was also investigated. Here the point is that VEGF was not measured during the trial. VEGF at tumor site would have been the closest surrogate for action on the tumor growth via vascularisation process. The bound aflibercept is then the closest surrogate to link to the effect on tumor cell growth. Exposure of aflibercept driven by the bound aflibercept concentrations is therefore more likely to be mechanistically correct.

The developed TGI model provided good parameter estimates and good individual fits for various tumor size profiles in both the reference and the aflibercept groups. The VPC suggested a good predictive performance of this TGI model at early times (i.e up to 45 weeks). Therefore, good tumor size metrics can be derived from this model and can be used in survival model to predict the efficacy of the therapy [3,16]. The common metrics is early measure of tumor size ratio (change in tumor size from the baseline), for example at weeks 6 to 8. More recently, the time to tumor regrowth has been shown to be a good metric to capture

the effect of bevacizumab and predict survival in first-line colorectal cancer patients [7]. In addition, by incorporating the PK of aflibercept and dose history of FOLFIRI, this model will be helpful to simulate efficacy rates at different dosing regimens. However, additional analyses should be further performed to improve this model, such as the covariate analysis to reduce the high variability on parameters and the joint modeling taking into account the large quantity of drop-out data which results in observed tendency in NPDE plots at lately times and may give biased estimates of parameters.

In summary, the tumor response and their relationship to the PK of VEGF-bound aflibercept and dosing history of FOLFIRI was successfully incorporated in a TGI model. This model described reasonably well the early measurement of tumor size, but further improvement will be necessary before using it to determine rational dosing strategies by simulation.

References:

1. Tham LS, Wang L, Soo RA, Lee SC, Lee HS, Yong WP, Goh BC, Holford NH (2008) A pharmacodynamic model for the time course of tumor shrinkage by gemcitabine + carboplatin in non-small cell lung cancer patients. *Clinical Cancer Research* 14 (13):4213-4218
2. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, Gobburu J (2009) Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clinical Pharmacology and Therapeutics* 86 (2):167-174
3. Bruno R, Lu JF, Sun YN, Claret L (2011) A modeling and simulation framework to support early clinical drug development decisions in oncology. *Journal of Clinical Pharmacology* 51 (1):6-8
4. Ribba B, Kaloshi G, Peyre M, Ricard D, Calvez V, Tod M, Cajavec-Bernard B, Idhahbi A, Psimaras D, Dainese L, Pallud J, Cartalat-Carel S, Delattre JY, Honnorat J, Grenier E, Ducray F (2012) A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clinical Cancer Research* 18 (18):5071-5080
5. Claret L, Girard P, Hoff PM, Van Cutsem E, Zuideveld KP, Jorga K, Fagerberg J, Bruno R (2009) Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *Journal of Clinical Oncology* 27 (25):4103-4108
6. Lu JF, Claret L, Sutjandra L, Kuchimanchi M, Melara R, Bruno R, Sun YN (2010) Population pharmacokinetic/pharmacodynamic modeling for the time course of tumor shrinkage by motesanib in thyroid cancer patients. *Cancer Chemotherapy and Pharmacology* 66 (6):1151-1158
7. Claret L, Gupta M, Joshi A, Sarapa N, He J, Powell B, Bruno R (2012) Evaluation of tumor-size response metrics to predict survival and progression free survival in first-line metastatic colorectal cancer. *PAGE* 21 (2012) abstr 2328 [www.page-meeting.org/?abstract=2328]
8. Chu QS (2009) Aflibercept (AVE0005): an alternative strategy for inhibiting tumour angiogenesis by vascular endothelial growth factors. *Expert Opinion on Biological Therapy* 9 (2):263-271

9. Gaya A, Tse V (2012) A preclinical and clinical review of aflibercept for the management of cancer. *Cancer Treatment Reviews* 38 (5):484-493
10. Van Cutsem E, Tabernero J, Lakomy R, Prenen H, Prausova J, Macarulla T, Ruff P, van Hazel GA, Moiseyenko V, Ferry D, McKendrick J, Polikoff J, Tellier A, Castan R, Allegra C (2012) Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *Journal of Clinical Oncology* 30 (28):3499-3506
11. Thai HT, Veyrat-Follet C, Vivier N, Dubruc C, Sanderink G, Mentre F, Comets E (2011) A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects. *British Journal of Clinical Pharmacology* 72 (3):402-414
12. Thai HT, Veyrat-Follet C, Mentre F, Comets E Population pharmacokinetic analysis of free and bound aflibercept in patients with advanced solid tumors. *World Conference on Pharmacometrics* (2012), Seoul, Korea.
13. Jacqmin P, Snoeck E, van Schaick EA, Gieschke R, Pillai P, Steimer JL, Girard P (2007) Modelling response time profiles in the absence of drug concentrations: definition and performance evaluation of the K-PD model. *Journal of Pharmacokinetics and Pharmacodynamics* 34 (1):57-85
14. Brendel K, Comets E, Laffont C, Laveille C, Mentre F (2006) Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharmaceutical Research* 23 (9):2036-2049
15. Comets E, Brendel K, Mentre F (2008) Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Computer Methods and Programs in Biomedicine* 90 (2):154-166
16. Bruno R, Claret L (2009) On the use of change in tumor size to predict survival in clinical oncology studies: toward a new paradigm to design and evaluate phase II studies. *Clinical Pharmacology and Therapeutics* 86 (2):136-138

Chapitre 3

L'approche bootstrap pour l'estimation d'incertitude des paramètres dans le cadre des modèles non linéaires à effets mixtes

3.1 Développement et comparaison de différentes approches bootstrap dans les modèles linéaires à effets mixtes

3.1.1 Résumé

Dans le développement des modèles, par exemple les modèles mécanistiques présentés dans le chapitre précédent, l'incertitude liée à l'estimation des paramètres peut être biaisé et parfois n'est pas obtenue. Dans ce cas, nous pouvons utiliser l'approche bootstrap qui consiste à rééchantillonner avec remise les données observées de façon répétée (EFRON, 1979). À ce jour, la plupart des applications en PK/PD utilisent le bootstrap non-paramétrique par paires qui rééchantillonne les sujets comme une méthode de validation interne avancée (ETTE, 1997; PARKE et al., 1999). D'autres méthodes de bootstrap basées sur des résidus qui consistent à rééchantillonner à la fois des effets aléatoires et des résidus ont été proposées pour mieux prendre en compte la structure hiérarchique dans les données longitudinales avec des mesures répétées au sein d'un sujet (DAS et KRISHEN, 1999; HALIMI, 2005). Cependant, peu d'études ont été effectuées pour comparer les performances de différentes méthodes de bootstrap dans les MEM. Nous avons cherché à étudier et à proposer des méthodes de bootstrap appropriées dans les MEM. Dans ce travail, nous avons tout d'abord évalué leur performance dans les modèles linéaires à effets mixtes (MLEM) avec l'erreur homoscédastique par l'intermédiaire d'études de simulation basé sur un sous-ensemble des données réelles décrivant l'évolution naturelle de la maladie de Parkinson (HOLFORD et al., 2006).

Différentes approches bootstrap qui tiennent compte de deux niveaux de variabilité (inter-sujet et résiduelle) dans MEM ont été mises en œuvre dans R (version 2.14.1) et

la fonction lme a été utilisée pour ajuster les données. Le rééchantillonnage du BSV peut être réalisé par sujets ou effets aléatoires. Le rééchantillonnage du RUV peut être réalisé en rééchantillonnant des résidus obtenus à partir de l'ensemble des sujets ou des résidus au sein de chaque sujet. Les effets aléatoires et les résidus peuvent être rééchantillonnés dans les distributions empiriques (l'approche non paramétrique) ou simulés à partir des distributions estimées (l'approche paramétrique). Ces méthodes ont également été comparées au bootstrap des résidus seul et au bootstrap par paires. Nous avons étudié les corrections apportées aux effets aléatoires et aux résidus pour tenir compte de la sous-estimation de la variance. Nous avons également comparé deux méthodes d'estimation, le maximum de vraisemblance restreint (REML) qui est principalement utilisé dans les MLEM et le maximum de vraisemblance (ML) qui est souvent utilisé dans les MNLEM. La performance des méthodes de bootstrap a été évaluée dans des conditions différentes : un protocole riche ($N=100$ sujets et $n=7$ échantillons), un protocole allégé ($N=30/n=3$) et un protocole de large erreur ($N=100/n=7$). Pour chaque modèle, 1000 répliques ont été simulées, et 1000 échantillons de bootstrap par réplique ont été générés pour chaque méthode de bootstrap. Les performances des méthodes de bootstrap ont été comparées en terme de biais de paramètres, de biais d'erreurs standard (SE) et de taux de recouvrement des intervalles de confiance à 95%. Elles ont également été comparées à celle de la méthode asymptotique.

Comme attendu, notre étude de simulation a montré une bonne performance de l'approche asymptotique. Elle a aussi montré une bonne performance du bootstrap par paires et des bootstraps non paramétrique/paramétrique des deux effets aléatoires et des résidus. D'autre part, les méthodes de bootstrap qui rééchantillonnent seulement les résidus et le bootstrap par paires couplé avec le bootstrap des résidus avaient de mauvaises performances. En appliquant les méthodes avec de bonnes performances pour l'ensemble des données réelles, nous avons constaté des différences entre elles, qui n'ont pas été observées lors de l'étude de simulation. Ces différences ont été liées à la distribution non normale des résidus et au protocole déséquilibré dans les données réelles. Concernant les méthodes d'estimation, REML et ML ont fourni des estimations bootstrap d'incertitude similaires, mais il y avait des biais légers et des taux de recouvrement plus pauvres pour les paramètres de variance avec ML dans le protocole allégé.

Ce travail a fait l'objet d'un article publié dans la revue *Pharmaceutical Statistics*.

3.1.2 Article 3 (publié)

A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models[†]

Hoai-Thu Thai,^{a*} France Mentré,^a Nicholas H. G. Holford,^b Christine Veyrat-Follet,^c and Emmanuelle Comets^a

A version of the nonparametric bootstrap, which resamples the entire subjects from original data, called the case bootstrap, has been increasingly used for estimating uncertainty of parameters in mixed-effects models. It is usually applied to obtain more robust estimates of the parameters and more realistic confidence intervals (CIs). Alternative bootstrap methods, such as residual bootstrap and parametric bootstrap that resample both random effects and residuals, have been proposed to better take into account the hierarchical structure of multi-level and longitudinal data. However, few studies have been performed to compare these different approaches. In this study, we used simulation to evaluate bootstrap methods proposed for linear mixed-effects models. We also compared the results obtained by maximum likelihood (ML) and restricted maximum likelihood (REML). Our simulation studies evidenced the good performance of the case bootstrap as well as the bootstraps of both random effects and residuals. On the other hand, the bootstrap methods that resample only the residuals and the bootstraps combining case and residuals performed poorly. REML and ML provided similar bootstrap estimates of uncertainty, but there was slightly more bias and poorer coverage rate for variance parameters with ML in the sparse design. We applied the proposed methods to a real dataset from a study investigating the natural evolution of Parkinson's disease and were able to confirm that the methods provide plausible estimates of uncertainty. Given that most real-life datasets tend to exhibit heterogeneity in sampling schedules, the residual bootstraps would be expected to perform better than the case bootstrap. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: bootstrap; longitudinal data; Parkinson's disease; linear mixed-effects models; R

1. INTRODUCTION

Mixed-effects models are commonly used to analyze longitudinal data that consist of repeated measures from individuals through time [1]. They play an important role in medical research, particularly in clinical trials. These models incorporate the fixed effects, which are parameters representing effects in the entire population, and random effects, which are associated with individuals sampled from a population [2]. The parameters of a model are estimated by maximum likelihood (ML) or restricted maximum likelihood (REML) method. In linear mixed-effects models (LMEMs), REML is often preferred to ML estimation because it takes into account the loss of the degrees of freedom involved in estimating the fixed effects, resulting in unbiased estimates of variance components in many situations [2, 3].

The standard errors (SEs) of parameter estimates are obtained asymptotically from the inverse of the Fisher information matrix [2, 3]. The aforementioned estimates of SE might be biased when the asymptotic approximation is incorrect, for example, when the sample size is small. Sometimes, they cannot be obtained when the model is complex or the design is too sparse. Bootstrap methods represent an alternative approach for estimating the SE of parameters, as well as to provide a CI without assuming it is symmetrical. It was first introduced by Efron (1979) for independent and identically distributed observations. The principal idea of bootstrap is to resample the observed data repeatedly to

create datasets similar to the original dataset and then fit them to construct the distribution of an estimator or a statistic of interest [4, 5]. Four main bootstrap approaches have been proposed for simple linear regression: case bootstrap, residual bootstrap, parametric bootstrap, and wild bootstrap [6–9]. The case bootstrap is the most simple and intuitive form that consists in resampling the entire vector of observations with replacement. The residual bootstrap resamples the residuals after model fitting. The parametric bootstrap adopts the principle of residual bootstrap, but instead of directly resampling observed residuals, we simulate the residuals from the estimated distribution, for example, the normal distribution, whose parameters are estimated using the original data. The wild bootstrap consists in resampling the residuals from an external distribution satisfying certain specifications.

[†]Supporting information may be found in the online version of this article.

^aUniv Paris Diderot, Sorbonne Paris Cité, UMR 738, F-75018 Paris, France; INSERM, UMR 738, F-75018 Paris, France

^bDepartment of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand

^cDrug Disposition Department, Sanofi, Paris, France

*Correspondence to: Hoai-Thu Thai, UMR738 INSERM, University Paris Diderot, 75018 Paris, France.
E-mail: hoai-thu.thai@inserm.fr

The main concern when bootstrapping is how to generate a bootstrap distribution close to the true distribution of the original sample. To do that, the bootstrap resampling should appropriately mimic the 'true' data generating process that produced the 'true' dataset [10–12]. In the context of repeated measurement data and mixed-effects modeling, the bootstrap should therefore respect the true data generating process with the repeated measures within a subject and handle two levels of variability: between-subject variability (BSV) and residual variability (RUV). The classical bootstrap methods developed in simple linear regression should be modified to take into account the characteristics of mixed-effects models [13]. Resampling random effects may be coupled with resampling residuals [10, 13–15]. The case bootstrap can be combined with the residual bootstrap [8]. The performance of these approaches are, however, not well studied.

In this paper, we extend different bootstrap approaches that can be applied to LMES settings. The detail of bootstrap methods is described in Section 2. The simulation settings are described in Section 3. The results of the simulation studies and the application of the bootstraps to real data collected in a study in Parkinson's disease are described in Section 4. The discussion of the study is given in Section 5.

2. METHODS

2.1. Statistical models

Let \mathbf{Y} be the response variable. N denotes the number of subjects. Let y_{ij} denote the observation j of \mathbf{Y} in subject i , whereas $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ regroups the $(n_i \times 1)$ vector of measurements in this subject. Let $n_{\text{tot}} = \sum_{i=1}^N n_i$ denote the total number of observations. We use the following LMES [16]:

$$\begin{cases} \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\eta}_i \sim N(0, \Omega) \\ \boldsymbol{\epsilon}_i \sim N(0, \sigma^2 \mathbf{I}_{n_i}) \end{cases} \quad (1)$$

where \mathbf{X}_i and \mathbf{Z}_i are $(n_i \times p)$ and $(n_i \times q)$ design matrices, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector containing the fixed effects, $\boldsymbol{\eta}_i$ is the $(q \times 1)$ vector containing the random effects, and $\boldsymbol{\epsilon}_i$ is the $(n_i \times 1)$ vector of residual components. Ω is the general $(q \times q)$ covariance matrix with (i, j) element $\omega_{ij} = \omega_{ji}$, and $\sigma^2 \mathbf{I}_{n_i}$ is the $(n_i \times n_i)$ covariance matrix for residual errors in subject i where σ^2 is the error variance and \mathbf{I}_{n_i} is the $(n_i \times n_i)$ identity matrix. The random effects $\boldsymbol{\eta}_i$ are assumed to be normally distributed with mean 0 and covariance matrix Ω , and the residual errors $\boldsymbol{\epsilon}_i$ are assumed to be normally distributed with mean 0 and variance $\sigma^2 \mathbf{I}_{n_i}$. The random effects $\boldsymbol{\eta}_i$ and the residual errors $\boldsymbol{\epsilon}_i$ are assumed to be independent for different subjects and to be independent of each other for the same subject.

Conditional on the random effects $\boldsymbol{\eta}_i$, the response \mathbf{Y}_i in subject i is normally distributed with mean vector $\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\eta}_i$ and with covariance matrix $\sigma^2 \mathbf{I}_{n_i}$.

2.2. Estimation methods

The parameters of LMESs can be estimated in the framework of ML by two general methods: ML or REML. Let $\boldsymbol{\alpha}$ denote the vector of all variance components of $\mathbf{V}_i = \mathbf{Z}_i \Omega \mathbf{Z}_i' + \Sigma_i$; it means $\boldsymbol{\alpha}$ consists of the $q(q+1)/2$ different elements in Ω (or q elements if Ω is diagonal) and of all parameters in Σ_i . Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ be the $(s \times 1)$ vector of all parameters in the marginal model for \mathbf{Y}_i . The parameter estimates are obtained by maximizing the marginal likelihood

function with respect to $\boldsymbol{\theta}$. The ML likelihood function is defined as follows:

$$L_{\text{ML}}(\boldsymbol{\theta}) = \prod_{i=1}^N \left\{ (2\pi)^{-n_i/2} |\mathbf{V}_i|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right) \right\} \quad (2)$$

The REML likelihood function is derived from $L_{\text{ML}}(\boldsymbol{\theta})$ to correct the loss of the degrees of freedom involved in estimating the fixed effects:

$$L_{\text{REML}}(\boldsymbol{\theta}) = \left| \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right|^{-1/2} L_{\text{ML}}(\boldsymbol{\theta}) \quad (3)$$

In this study, we used REML as the estimation method. However, we also compared the results of REML with those of ML and presented them in the Appendix (available online as supplementary material).

When $\boldsymbol{\alpha}$ is known, the maximum likelihood estimator of $\boldsymbol{\beta}$, obtained from maximizing $L_{\text{ML}}(\boldsymbol{\theta})$, conditional on $\boldsymbol{\alpha}$, is given by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \quad (4)$$

When $\boldsymbol{\alpha}$ is unknown, but an estimate $\hat{\boldsymbol{\alpha}}$ is available, we can set $\mathbf{V}_i = \hat{\mathbf{V}}_i$ and estimate $\boldsymbol{\beta}$ by using Equation (4).

Estimates of the $\boldsymbol{\eta}_i$ can be obtained as the mean of the posterior distribution of $\boldsymbol{\eta}_i$ (empirical Bayes estimates):

$$\hat{\boldsymbol{\eta}}_i(\boldsymbol{\theta}) = E[\boldsymbol{\eta}_i | \mathbf{y}_i] = \int \boldsymbol{\eta}_i f(\boldsymbol{\eta}_i | \mathbf{y}_i) d\boldsymbol{\eta}_i = \Omega \mathbf{Z}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \quad (5)$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is asymptotically normally distributed with mean $\boldsymbol{\theta}$ and asymptotic covariance matrix given by the inverse of the Fisher information matrix M_F . The asymptotic SEs of parameters are then estimated as the square root of the diagonal element of the estimated covariance matrix.

2.3. Bootstrap methods

The principle of the bootstrap is to repeatedly generate pseudo-samples distributed according to the same distribution as the original sample. The unknown original distribution may be replaced by the empirical distribution of the sample, which is known as the nonparametric bootstrap [17]. The bootstrap exists in another version called the parametric bootstrap [7, 9]. In this version, the underlying distribution F is estimated from the data by a parametric model, for instance, normal distribution and bootstrap samples are generated by simulating within this distribution rather than from the empirical distribution as performed in the nonparametric version. The resampling can be carried out with an independent distribution; this procedure is called external bootstrap or wild bootstrap. This approach was proposed by Wu to deal with heteroscedasticity [18]. We have chosen to deal with the simpler case of homoscedasticity and therefore have not investigated the wild bootstrap.

Let B be the number of bootstrap samples to be drawn from the original dataset; a general bootstrap algorithm is as follows:

- (1) Generate a bootstrap sample by resampling from the data and/or from the estimated model
- (2) Obtain the estimates for all parameters of the model for the bootstrap sample

- (3) Repeat steps 1 and 2 B times to obtain the bootstrap distribution of parameter estimates and then compute the mean, standard deviation, and 95% CI of this distribution.

Let $\hat{\theta}_b^*$ be the parameter estimated for the b th bootstrap sample. Given a data set, the expected value of the bootstrap estimator over the bootstrap distribution is calculated as the average of the parameter estimates from the B bootstrap samples:

$$\hat{\theta}_B = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^*) \quad (6)$$

The bootstrap SE is obtained as the sample standard deviation of the $\hat{\theta}_b^*$:

$$\widehat{SE}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}_B)^2} \quad (7)$$

A 95% bootstrap CI can be constructed by calculating the 2.5th and 97.5th percentile of bootstrap distribution:

$$\hat{\theta}_{(\alpha \cdot B)}^* \leq \theta \leq \hat{\theta}_{((1-\alpha) \cdot B)}^* \quad (8)$$

where $\alpha = 0.025$. An alternative approach is to use a normal approximation to construct a bootstrap CI, using the estimated \widehat{SE}_B :

$$\hat{\theta}_B - \widehat{SE}_B \cdot z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_B + \widehat{SE}_B \cdot z_{1-\alpha/2} \quad (9)$$

$z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution, equal to 1.96 with $\alpha = 0.05$. However, it is preferable to use the bootstrap percentile CI when bootstrapping [7, 19].

The detailed algorithms of bootstrap methods to obtain a bootstrap sample (bootstrap generating process) are presented next in two separated groups: nonparametric and parametric bootstrap methods.

2.3.1. Nonparametric bootstrap. Nonparametric case bootstrap ($B_{\text{case,none}}$). This method consists of resampling with replacement the entire subjects, that is, the joint vector of design variables and corresponding responses ($\mathbf{X}_i, \mathbf{Z}_i, \mathbf{y}_i$) from the original data before modeling. It is also called the *paired bootstrap*. This procedure omits the second step of resampling the observations inside each subject. However, it is the most obvious way to do bootstrapping and makes no assumptions on the model.

Nonparametric case bootstrap coupled with global/individual residual bootstrap ($B_{\text{case,GR}}$ or $B_{\text{case,IR}}$). This method resamples first the entire subjects with replacement. The individual residuals are then resampled with replacement globally from the residual distribution of the original simulated dataset or individually from the residual distribution of new subjects obtained after bootstrapping in the first step. The bootstrap sample is obtained as follows:

- (1) Fit the model to the data and then calculate the residuals $\hat{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\eta}_i$
- (2) Draw N entire subjects $\{(\mathbf{X}_i^*, \mathbf{Z}_i^*, \mathbf{y}_i^*)\}$ with replacement from $\{(\mathbf{X}_i, \mathbf{Z}_i, \mathbf{y}_i)\}$ in the original data and keep their predictions from model fitting $\mathbf{X}_i^* \hat{\beta} + \mathbf{Z}_i^* \hat{\eta}_i^*$ and their corresponding residuals $\hat{\epsilon}_i^* = \mathbf{y}_i^* - \mathbf{X}_i^* \hat{\beta} - \mathbf{Z}_i^* \hat{\eta}_i^*$. The new subject has n_i^* observations.

- (3) Draw the residuals with replacement globally from all residuals of the original data or individually from each new subject

- (a) Global residual resampling: draw a sample $\{\epsilon^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size $n_{\text{tot}}^* = \sum_{i=1}^N n_i^*$ with replacement globally from $\{\hat{\epsilon}\} = \{\hat{\epsilon}_{ij}\}_{i=1, \dots, N; j=1, \dots, n_i}$ by assigning an equal probability $\frac{1}{n_{\text{tot}}}$ to each value of the n_{tot} residuals (note that, n_{tot}^* may be different with n_{tot})
- (b) Individual residual resampling: draw individually N samples $\{\epsilon_i^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size n_i^* with replacement from samples $\hat{\epsilon}_i^* = \{\hat{\epsilon}_{ij}\}_{j=1, \dots, n_i^*}$ by assigning an equal probability $\frac{1}{n_i^*}$ to each residual of new subject i

- (4) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i^* \hat{\beta} + \mathbf{Z}_i^* \eta_i^* + \epsilon_i^*$

Here, and in the following, we note the vector of estimated residuals for all subjects as $\hat{\epsilon} = \{\hat{\epsilon}_{ij}\}_{i=1, \dots, N; j=1, \dots, n_i}$ and the vector of estimated residuals in subject i as $\hat{\epsilon}_i = \{\hat{\epsilon}_{ij}\}_{j=1, \dots, n_i}$.

Nonparametric random effects bootstrap coupled with global/individual residual bootstrap ($B_{\eta, \text{GR}}$ or $B_{\eta, \text{IR}}$). This method consists of resampling with replacement the random effects obtained after model fitting, as well as the residuals globally or individually. The bootstrap sample is obtained as follows:

- (1) Fit the model to the data and then estimate the random effects $\hat{\eta}_i$ and the residuals $\hat{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\eta}_i$
- (2) Draw a sample $\{\eta_i^*\}$ of size N with replacement from $\{\hat{\eta}_i\}$ by assigning an equal probability $\frac{1}{N}$ to each value
- (3) Draw a sample $\{\epsilon^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size n_{tot} with replacement globally from $\{\hat{\epsilon}\}$ or draw individually N samples $\{\epsilon_i^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size n_i with replacement from $\{\hat{\epsilon}_i\}$
- (4) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i \hat{\beta} + \mathbf{Z}_i \eta_i^* + \epsilon_i^*$

Nonparametric global/individual residual bootstrap ($B_{\text{none,GR}}$ or $B_{\text{none,IR}}$). For the sake of completeness, we also implemented a bootstrap where only RUV is resampled. These procedures do not resample the BSV (which remains in the model through the estimated random effects $\hat{\eta}_i$). The bootstrap sample is obtained as follows:

- (1) Fit the model to the data and then calculate the residuals $\hat{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\eta}_i$
- (2) Draw a sample $\{\epsilon^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size n_{tot} with replacement globally from $\{\hat{\epsilon}\}$ or draw individually N samples $\{\epsilon_i^*\} = \{\hat{\epsilon}_{ij^*}\}$ of size n_i with replacement from $\{\hat{\epsilon}_i\}$
- (3) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i \hat{\beta} + \mathbf{Z}_i \hat{\eta}_i + \epsilon_i^*$

The nonparametric bootstrap methods that resample the random effects or the residuals depend on the structural model to calculate the raw random effects or residuals. However, they do not require particular assumptions on their distributions.

2.3.2. Parametric bootstrap. The parametric bootstrap requires the strongest assumptions as it depends both on the model and the distributions of parameters and errors.

Case bootstrap coupled with parametric residual bootstrap ($B_{\text{case,PR}}$). Similar to $B_{\text{case,GR}}$, this method resamples firstly the subjects and then resamples the residuals by simulating from the estimated distribution. This method combines elements of both the nonparametric bootstrap (case bootstrap in the first step) and the parametric bootstrap (residual bootstrap in the second step).

However, to simplify the classification, we keep it in the group of parametric bootstraps. The bootstrap sample is obtained as follows:

- (1) Fit the model to the data
- (2) Draw N entire subjects $\{(\mathbf{X}_i^*, \mathbf{Z}_i^*, \mathbf{y}_i^*)\}$ with replacement from $\{(\mathbf{X}_i, \mathbf{Z}_i, \mathbf{y}_i)\}$ in the original data and keep their predictions from model fitting $\mathbf{X}_i^* \hat{\beta} + \mathbf{Z}_i^* \hat{\eta}_i^*$
- (3) Draw N samples $\{\epsilon_i^*\}$ of size n_i^* from a normal distribution with mean 0 and covariance matrix $\hat{\sigma}^2 \mathbf{I}_{n_i}$
- (4) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i^* \hat{\beta} + \mathbf{Z}_i^* \hat{\eta}_i^* + \epsilon_i^*$

Parametric random effects bootstrap coupled with residual bootstrap ($B_{PR,PR}$). This methods resamples both random effects and residuals by simulating from estimated distribution after model fitting. The bootstrap sample is obtained as follows:

- (1) Fit the model to the data
- (2) Draw a sample $\{\eta_i^*\}$ of size N from a multivariate normal distribution with mean 0 and covariance matrix $\hat{\Sigma}$
- (3) Draw N samples $\{\epsilon_i^*\}$ of size n_i from a normal distribution with mean zero and covariance matrix $\hat{\sigma}^2 \mathbf{I}_{n_i}$
- (4) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i \hat{\beta} + \mathbf{Z}_i \eta_i^* + \epsilon_i^*$

Parametric residual bootstrap ($B_{none,PR}$). Again for exhaustiveness, we implemented a bootstrap that resamples only the residuals by simulating from the estimated distribution after model fitting. Similar to $B_{none,GR}$ or $B_{none,IR}$, this procedure omits the first step of resampling the BSV. The bootstrap sample is obtained as follows:

- (1) Fit the model to the data
- (2) Draw N samples $\{\epsilon_i^*\}$ of size n_i from a normal distribution with mean 0 and covariance matrix $\hat{\sigma}^2 \mathbf{I}_{n_i}$
- (3) Generate the bootstrap responses $\mathbf{y}_i^* = \mathbf{X}_i \hat{\beta} + \mathbf{Z}_i \hat{\eta}_i + \epsilon_i^*$

2.3.3. Transformation of random effects and residuals. Previous results in the literature show that the nonparametric bootstrap of the raw residuals obtained by model fitting usually yield downwardly biased variance parameter estimates [7, 8, 20]. In ordinary linear models, this underestimation is due to the difference between estimated and empirical variance of residuals [21]. That is why it is advisable to rescale the residuals so that they have the correct variance. Efron suggested to multiply centered raw residuals with the factor $\sqrt{(n-p)/n}$ where p is the number of parameters and n is the number of observations [22]. For the same reason, Davison *et al.* proposed to use the factor $\sqrt{1/(1-h_i)}$ where h_i is the i th diagonal element of the hat matrix that maps the vectors of observed values to the vector of fitted values [7]. In the mixed-effects models, the raw variance–covariance matrix is different from the ML estimate, as the raw random effects or residuals are ‘shrunk’ towards 0 [20]. Carpenter *et al.* proposed to take this into account by centering the random effects and residuals to resample from a distribution with mean 0 and multiplying them by the ratio between their corresponding estimated and empirical variance–covariance matrices to account for the variance underestimation [20, 23]. These corrections were used in our study.

Transforming random effects. The transformation of random effects was carried out in the following steps:

- (1) Center the raw estimated random effects: $\tilde{\eta}_i = \hat{\eta}_i - \bar{\eta}_i$
- (2) Calculate the ratio between the estimated and empirical variance–covariance matrix (A_η). Let $\hat{\Sigma}$ be the model estimated variance–covariance matrix of random effects and

Σ_{emp} denote the empirical variance–covariance matrix of the centered random effects $\tilde{\eta}_i$. The ratio matrix A_η is formed by using the Cholesky factors L_{est} and L_{emp} , which are the lower triangular matrix of the Cholesky decomposition of $\hat{\Sigma}$ and

Σ_{emp} , respectively: $A_\eta = (L_{\text{est}} \cdot L_{\text{emp}}^{-1})'$

- (3) Transform the centered random effects using the ratio A_η : $\hat{\eta}_i = \tilde{\eta}_i \times A_\eta$

Transforming residuals. The transformation of residuals was carried out globally for all residuals in the following steps:

- (1) Center the raw estimated residuals: $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} - \bar{\epsilon}_{ij}$
- (2) Calculate the ratio between the estimated and empirical variance–covariance matrix (A_σ). Let $\hat{\Sigma}$ be the model estimated variance–covariance matrix for residuals, which is assumed to be equal to $\hat{\sigma}^2 \mathbf{I}_{n_i}$ and Σ_{emp} denote the empirical variance–covariance matrix of centered residual $\tilde{\epsilon}_{ij}$. Because the residuals are assumed to be uncorrelated and to have equal variance, the ratio matrix A_σ is then simply the ratio between the square root of the model estimated residual variance $\hat{\sigma}^2$ and the empirical standard deviation of the centered residuals, respectively: $A_\sigma = \hat{\sigma} / \text{sd}(\tilde{\epsilon}_{ij})$.
- (3) Transform the centered residuals using the ratio A_σ : $\hat{\epsilon}'_{ij} = \tilde{\epsilon}_{ij} \times A_\sigma$

3. SIMULATION STUDIES

3.1. Motivating example

The motivating example for bootstrap evaluation was a disease progression model inspired from the model of Parkinson's disease developed by Holford *et al.* [24]. In that study, the subjects were initially randomized to treatment with placebo, deprenyl, tocopherol, or with both and, when clinical disability required, received one or more dopaminergic agents (levodopa, bromocriptine, or pergolide). The aim was to study the influence of various drugs on the changes in Unified Parkinson's Disease Rating Scale (UPDRS) over time. Several components describing the disease progression and the effect of treatment were developed. However, in this study, we are only interested in the linear part of the model, describing the natural evolution of Parkinson's disease using a random intercept and slope model as

$$y_{ij} = S_0 + \alpha \cdot t_{ij} + \eta_{S_0i} + \eta_{\alpha i} \cdot t_{ij} + \epsilon_{ij} \quad (10)$$

where y_{ij} is the UPDRS score (u) representing the state of disease at time t_{ij} that was considered to be continuous, S_0 (u) is the expected score at randomization, α (u/year) is the progression rate, η_{S_0i} (u/year) and $\eta_{\alpha i}$ (u/year) are the random effects of S_0 and α , respectively, and ϵ_{ij} is the residual errors. In the formulation of Equation (1), the design matrix $\mathbf{X}_i = \mathbf{Z}_i$ for subject i is a two-column matrix with a first column of 1s and a second column containing the n_i times t_{ij} for subject i .

For our simulations, we used the subset of patients who remained in the placebo group over the first 2 years. The UPDRS scores were measured at randomization and at regular (unscheduled) visits up to 2 years after entry to the study. This subset contains 109 subjects with an average of six observations per subject and a total of 822 observations. Figure 1 describes the evolution of UPDRS score over a two-year period.

In the paper of Holford *et al.*, the baseline S_0 was assumed to be log-normally distributed, and the progression rate α was

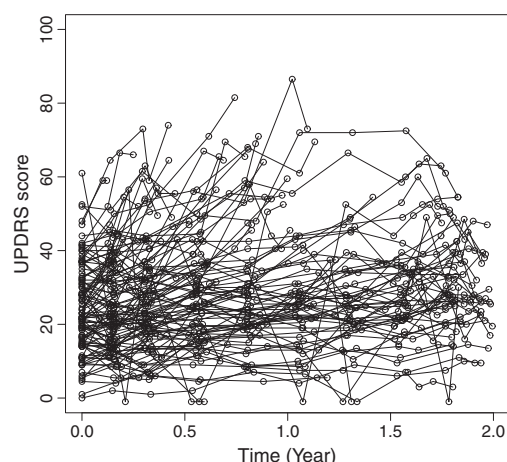


Figure 1. The evolution of Unified Parkinson's Disease Rating Scale (UPDRS) score over time in the real dataset used for the simulations, including 109 patients who remained in the placebo group over the first 2 years.

assumed to be normally distributed [24]. The variance of the random effects and the correlation ρ between S_0 and α was estimated. The residual unexplained variability was described by an additive model error with constant variance σ^2 . In the present study, both S_0 and α were assumed to be normally distributed. We estimated the parameters of the real dataset by ML, the same estimation method as used in the original publication [24], using the `lme` function in R. The detail of parameter estimates is given in Section 4.2.

3.2. Simulation settings

Our simulations were inspired from the original design described earlier using estimates from the real dataset.

Three designs were planned to evaluate the performance of bootstrap. For each design, the sampling times were similar for all subjects.

Rich design. We simulated $N = 100$ subjects with $n = 7$ observations per subject at 0, 0.17, 0.33, 0.5, 1, 1.5, and 2 years after being entered in the study.

Sparse design. We simulated $N = 30$ subjects with $n = 3$ observations per subject at 0, 0.17, and 2 years after being entered

in the study. This design is sparse with respect to estimation of variance parameters, including only 90 observations in total.

Large error design. We simulated $N = 100$ subjects with $n = 7$ observations per subject at 0, 0.17, 0.33, 0.5, 1, 1.5, and 2 years after being entered in the study. In this design, we modified the level of variability. The variability for random effects η_{S_0} and η_α was changed to $\omega_{S_0} = 11.09$ and $\omega_\alpha = 6.98$, respectively (equivalent to 50% of the corresponding fixed effects), and the standard deviation for the residual error was changed to $\sigma = 17.5$. We also removed the correlation ρ between S_0 and α in this design because convergence was obtained for only 78.3% simulated datasets with the presence of this correlation.

For each design, we simulated $K = 1000$ replications.

3.3. Software

The `lme` function in the `nlme` library in R was used to fit the data using REML as the estimation method. ML was also used to compare with the results of REML. For both methods, we fitted datasets with the initial values of variance parameters generated by the optimization procedure implemented in the `lme` function [25]. All the analysis and figures were carried out with R.

3.4. Evaluation of bootstrap methods

Table I presents all the bootstrap methods that we implemented and evaluated. The resampling of the BSV can be carried out by resampling subjects or random effects. The resampling of the RUV can be carried out by resampling of residuals obtained from all subjects, called global residuals, or resampling of residuals within each subject, called individual residuals. To compare the performance of these bootstraps to that of classical bootstraps, the residual bootstrap, which resamples only the RUV, and the case bootstrap, where the whole vector of observations, including both the BSV and RUV, is resampled, were also evaluated in our study. The bootstrap methods were classified as nonparametric and parametric methods. For the parametric approach, there is no difference between the global and the individual residual resampling because it is performed by simulation from the estimated distributions. In total, we had seven nonparametric methods and three parametric methods.

We drew $B = 1000$ bootstrap samples for each replication of simulated data and for each bootstrap method. The parameters were estimated by REML method using the `lme` function in R. For each method, we therefore performed 1 million fits (1000 simulated datasets \times 1000 bootstrap datasets). If the convergence was

Table I. Bootstrap methods that can be applied in mixed-effects models.

				Variability related to subject		
				None	Resample individuals	Resample random effects
Variability related to observation	Resample residuals	None	Original data	$B_{\text{case,none}}$		
		Globally	NP	$B_{\text{none,GR}}$	$B_{\text{case,GR}}$	$B_{\eta,\text{GR}}$
			P	$B_{\text{none,PR}}$	$B_{\text{case,PR}}$	$B_{P\eta,\text{PR}}$
		Individually	NP	$B_{\text{none,IR}}$	$B_{\text{case,IR}}$	$B_{\eta,\text{IR}}$

NP, nonparametric; P, parametric.

not obtained, the NA(not applicable) was recorded in the table of parameter estimates and excluded for further calculation.

For the k th simulated dataset and for a given bootstrap method, we computed the bootstrap parameter estimate $\hat{\theta}_B^k$ as in Equation (5), using 1000 bootstrap samples, as well as the bootstrap SE and CI, for each parameter θ . The relative bias (RBias) of bootstrap estimate was obtained by comparing the bootstrap estimate $\hat{\theta}_B^k$ and the asymptotic estimate $\hat{\theta}_k$ as follows:

$$\text{RBias}(\hat{\theta}_B) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{\theta}_B^k - \hat{\theta}_k}{\hat{\theta}_k} \times 100 \right) \quad (11)$$

The average bootstrap SE was obtained by averaging the SE from Equation (6) over the $K = 1000$ datasets. The true SE is unknown, but we can have an empirical estimate as the standard deviation of the differences between the estimate of the parameter in the K datasets and the true value (θ_0):

$$\text{SE}_{\text{empirical}}(\hat{\theta}) = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_k - \theta_0)^2} \quad (12)$$

The relative bias on bootstrap SE was then obtained by comparing the average bootstrap SE to this empirical SE:

$$\text{RBias}(\text{SE}(\hat{\theta}_B)) = \frac{\frac{1}{K} \sum_{k=1}^K \text{SE}_B(\hat{\theta}_B^k) - \text{SE}_{\text{empirical}}(\hat{\theta})}{\text{SE}_{\text{empirical}}(\hat{\theta})} \times 100 \quad (13)$$

The coverage rate of the 95% bootstrap CI was defined as the percentage of the $K = 1000$ datasets in which the bootstrap CI contains the true value of the parameter.

The bootstrap approaches were compared in terms of the Rbias on the bootstrap parameter estimates, the Rbias on SE, and the coverage rate of the 95% CI of all parameter estimates from one million bootstrap samples.

The performance of the bootstrap methods were also compared with the performance of the asymptotic method. It is worth noting that in the simulation studies, random variables were simulated according to normal distributions, which may have contributed to the good performance of the asymptotic method. The

relative bias of asymptotic estimate was obtained by comparing the asymptotic estimate $\hat{\theta}_k$ and the true value θ_0 as follows:

$$\text{RBias}(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{\theta}_k - \theta_0}{\theta_0} \times 100 \right) \quad (14)$$

The relative bias of asymptotic SEs given by the software (obtained as the inverse of the Fisher information matrix) was defined in the same way as Equation (13) but with respect to $\hat{\theta}_k$ instead of $\hat{\theta}_B^k$. The coverage rate of the 95% asymptotic CI was defined as the percentage of datasets in which the asymptotic CI contains the true value of the parameter.

The asymptotic and bootstrap parameters estimates and their SE were defined as unbiased when relative bias was within $\pm 5\%$, moderately biased (relative bias from $\pm 5\%$ to $\pm 10\%$) or strongly biased (relative bias $> \pm 10\%$). The coverage rate of the 95% CI was considered to be good (from 90% to 100%), low (from 80% to 90%), or poor ($< 80\%$). A good bootstrap was defined as a method providing unbiased estimates for the parameters and their corresponding SE and ensuring a good coverage rate of the 95% CI.

3.5. Application to real data

All the bootstrap methods with good performance evaluated in the simulations studies were applied to the real data by drawing $B = 1000$ bootstrap samples for each method. The bootstrap parameter estimates and bootstrap SE were compared with each other and compared with the parameter estimates and their SEs obtained by the asymptotic approach.

4. RESULTS

4.1. Simulation studies

Examples of simulated data for each given design are illustrated in Figure 2. Our simulations gave some negative values for the observations because of the homoscedastic error model. These values were kept as is, because the purpose of the simulations was not to provide a realistic simulation of the trial but only to

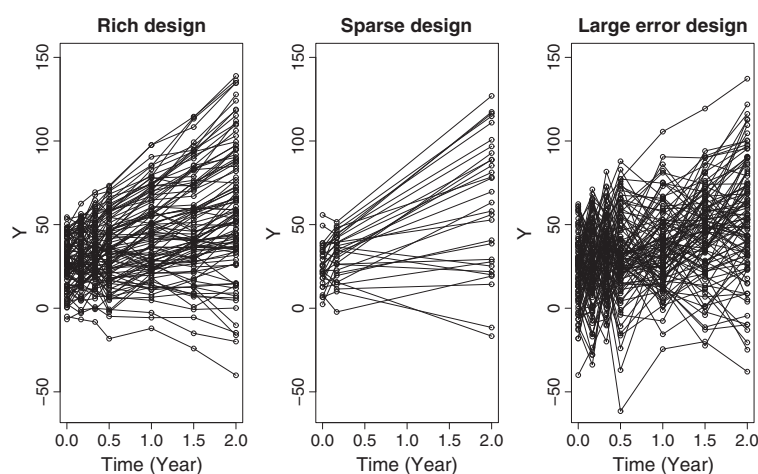


Figure 2. Examples of simulated data for each design.

Table II. Relative bias of parameter estimates by REML and their standard errors (SE) for the asymptotic method and the bootstrap methods in the three studied designs.

Design	Method	Relative bias of parameters (%)						Relative bias of SE (%)					
		S_0	α	ω_{S_0}	ω_α	ρ	σ	S_0	α	ω_{S_0}	ω_α	ρ	σ
Rich	Asymptotic	0.05	0.19	-0.13	-0.03	-0.27	0.03	2.79	1.62	-2.72	-1.02	4.99	-0.92
	$B_{\text{case,none}}$	0.00	0.01	-0.85	-0.81	0.15	-0.05	2.31	1.17	-5.47	-2.97	5.04	-1.99
	$B_{\text{none,GR}}$	0.00	0.00	-3.63	-2.53	13.48	-0.12	-70.16	-75.26	-60.48	-66.13	-41.84	0.14
	$B_{\text{none,IR}}$	0.00	0.00	0.51	-2.53	9.73	-0.61	-70.28	-75.37	-60.26	-65.89	-39.25	-2.66
	$B_{\text{none,PR}}$	0.04	0.19	-3.72	-2.53	12.97	-0.02	-70.15	-75.22	-60.45	-66.06	-41.80	-1.04
	$B_{\text{case,GR}}$	0.00	0.01	-4.34	-3.25	13.23	-0.13	-1.05	-1.07	-7.75	-4.89	-7.49	0.24
	$B_{\text{case,IR}}$	0.05	0.21	-0.46	-3.32	9.43	-0.67	6.97	1.87	3.30	0.92	14.82	65.57
	$B_{\text{case,PR}}$	0.04	0.18	-4.43	-3.26	12.70	-0.01	-1.19	-1.28	-7.54	-4.87	-7.34	-1.01
	$B_{\eta,\text{GR}}$	0.00	-0.01	-0.80	-0.78	-0.03	-0.12	2.34	1.10	-4.88	-2.65	5.59	0.07
	$B_{\eta,\text{IR}}$	0.00	0.00	-0.65	-0.77	-0.17	-0.62	2.26	1.00	-4.68	-2.68	6.28	-2.68
	$B_{P\eta,\text{PR}}$	0.04	0.19	-0.44	-0.32	-0.30	-0.02	2.75	1.57	-2.29	-0.64	5.94	-0.93
Sparse	Asymptotic	0.08	0.48	-0.92	-0.79	0.60	-0.99	2.96	0.04	-1.30	-2.51	-3.86	0.74
	$B_{\text{case,none}}$	-0.01	-0.02	-2.67	-2.63	1.24	-1.29	1.09	-1.85	-8.96	-9.77	-6.88	-8.12
	$B_{\text{none,GR}}$	0.00	0.00	-5.09	-3.10	19.24	-2.82	-62.90	-72.23	-50.93	-62.00	-40.45	4.04
	$B_{\text{none,IR}}$	0.00	0.00	6.83	-3.23	10.52	-6.46	-64.31	-73.39	-50.54	-63.16	-32.37	11.24
	$B_{\text{none,PR}}$	0.00	-0.01	-5.08	-3.11	19.25	-2.04	-62.72	-72.11	-50.49	-61.97	-40.03	-4.61
	$B_{\text{case,GR}}$	0.02	0.03	-7.33	-5.44	17.86	-2.98	-3.14	-4.43	-10.84	-10.95	-16.09	3.38
	$B_{\text{case,IR}}$	0.09	0.49	3.05	-6.19	5.37	-10.61	12.39	-1.15	6.84	-5.09	10.56	59.45
	$B_{\text{case,PR}}$	0.00	0.02	-7.27	-5.44	17.89	-2.24	-3.10	-4.32	-10.62	-10.83	-15.62	-5.00
	$B_{\eta,\text{GR}}$	0.02	0.02	-2.41	-2.46	0.37	-2.16	1.21	-1.60	-6.94	-8.51	-4.38	6.33
	$B_{\eta,\text{IR}}$	-0.01	0.01	-0.94	-2.45	-1.26	-6.50	0.82	-1.95	-5.42	-8.94	0.29	10.81
	$B_{P\eta,\text{PR}}$	0.00	0.03	-0.84	-0.88	-0.10	-1.34	2.97	0.12	-1.86	-2.92	-4.03	-2.11
Large error	Asymptotic	0.02	-0.06	-0.93	-1.24		0.02	-2.85	-2.05	-2.21	-3.57		1.12
	$B_{\text{case,none}}$	0.00	-0.01	-1.20	-3.58		-0.08	-3.39	-2.55	-3.53	1.06		-0.42
	$B_{\text{none,GR}}$	-0.01	0.01	-10.64	-24.75		-0.84	-37.47	-20.85	-29.94	5.66		-1.46
	$B_{\text{none,IR}}$	-0.01	0.01	18.63	-24.96		-2.70	-38.53	-22.32	-34.90	15.70		-5.98
	$B_{\text{none,PR}}$	0.00	0.00	-10.64	-24.74		-0.77	-37.44	-20.88	-29.85	5.73		-1.09
	$B_{\text{case,GR}}$	-0.01	-0.01	-11.37	-25.40		-0.83	-12.20	-12.45	-7.29	11.61		-1.29
	$B_{\text{case,IR}}$	0.01	0.00	16.85	-30.12		-2.96	19.51	15.75	19.38	61.88		60.31
	$B_{\text{case,PR}}$	0.00	0.00	-11.35	-25.43		-0.75	-12.11	-12.30	-7.22	11.80		-1.13
	$B_{\eta,\text{GR}}$	0.00	0.00	-1.07	-2.97		-0.12	-3.27	-2.13	-2.90	1.69		0.87
	$B_{\eta,\text{IR}}$	0.01	0.00	3.20	-3.02		-1.97	-3.92	-3.51	-1.81	3.35		-4.05
	$B_{P\eta,\text{PR}}$	0.00	0.01	-0.59	-2.46		-0.05	-2.87	-2.05	-1.78	2.13		1.01

Relative bias within $\pm 5\%$ is typeset in bold font.
REML, restricted maximum likelihood.

evaluate the bootstrap methods. For the same reason, we did not take into account other real-life factors such as drop-outs or missingness.

We present firstly the results of REML. Complete results for all bootstrap methods as well as the asymptotic method in the three evaluated designs are reported in Table II, for relative bias of the bootstrap estimates of parameters and their corresponding SEs, and Table III, for the coverage rate of the 95% CI. The correlation between S_0 and α was not estimated in the large error design for both simulated and bootstrap datasets to have better convergence rate and to keep the same model that we simulated. The transformation of random effects and residuals were carried out before resampling for the methods in which they were resampled nonparametrically. In the rich design, we found that the bootstrap methods that resample only the residuals ($B_{\text{none,GR}}$, $B_{\text{none,IR}}$,

$B_{\text{none,PR}}$) yielded higher bias for the correlation term ($>9.73\%$). The same was observed for the case bootstrap coupled with the residual bootstraps ($B_{\text{case,GR}}$, $B_{\text{case,IR}}$, $B_{\text{case,PR}}$). The case bootstrap and the bootstraps of both random effects and residuals ($B_{\eta,\text{GR}}$, $B_{\eta,\text{IR}}$, $B_{P\eta,\text{PR}}$) showed essentially no bias for all parameters. In terms of SE estimation, these four bootstrap methods estimated correctly all the SEs, as the estimates were very close to empirical SEs, whereas the residual-alone bootstraps greatly underestimated the SEs of all parameters, except for that of σ . On the contrary, the SE of σ was overestimated by the $B_{\text{case,IR}}$. The case bootstrap, coupled with global residual bootstrap ($B_{\text{case,GR}}$) or parametric bootstrap ($B_{\text{case,PR}}$), gave better estimates for SE of parameters than the residual-alone bootstraps, but they did not work as well as the case bootstrap and the bootstrap of random effects and residuals. In terms of coverage rate, the

Table III. Coverage rate of the 95% CI of parameters by REML obtained by the asymptotic method and the bootstrap methods in the three studied designs.

Design	Method	Coverage rate of the 95% CI of parameters					
		S_0	α	ω_{S_0}	ω_α	ρ	σ
Rich	Asymptotic	0.95	0.96	0.94	0.96	0.96	0.94
	$B_{\text{case,none}}$	0.94	0.96	0.92	0.94	0.95	0.94
	$B_{\text{none,GR}}$	0.44	0.36	0.50	0.44	0.46	0.95
	$B_{\text{none,IR}}$	0.43	0.36	0.56	0.44	0.60	0.93
	$B_{\text{none,PR}}$	0.43	0.36	0.50	0.44	0.45	0.94
	$B_{\text{case,GR}}$	0.94	0.95	0.85	0.89	0.72	0.94
	$B_{\text{case,IR}}$	0.96	0.96	0.95	0.91	0.91	1.00
	$B_{\text{case,PR}}$	0.94	0.95	0.85	0.88	0.72	0.94
	$B_{\eta,\text{GR}}$	0.94	0.95	0.92	0.95	0.95	0.95
	$B_{\eta,\text{IR}}$	0.95	0.95	0.92	0.94	0.96	0.94
	$B_{P\eta,\text{PR}}$	0.95	0.96	0.93	0.96	0.96	0.94
Sparse	Asymptotic	0.94	0.94	0.94	0.94	0.91	0.94
	$B_{\text{case,none}}$	0.94	0.93	0.89	0.89	0.92	0.90
	$B_{\text{none,GR}}$	0.53	0.39	0.60	0.51	0.66	0.95
	$B_{\text{none,IR}}$	0.51	0.38	0.65	0.48	0.84	0.92
	$B_{\text{none,PR}}$	0.53	0.40	0.61	0.50	0.66	0.93
	$B_{\text{case,GR}}$	0.93	0.93	0.83	0.86	0.84	0.95
	$B_{\text{case,IR}}$	0.96	0.94	0.98	0.88	0.99	0.99
	$B_{\text{case,PR}}$	0.93	0.93	0.83	0.86	0.85	0.93
	$B_{\eta,\text{GR}}$	0.94	0.93	0.91	0.89	0.94	0.95
	$B_{\eta,\text{IR}}$	0.94	0.94	0.93	0.89	0.96	0.92
	$B_{P\eta,\text{PR}}$	0.94	0.94	0.93	0.94	0.94	0.94
Large error	Asymptotic	0.93	0.95	0.95	0.98		0.95
	$B_{\text{case,none}}$	0.93	0.94	0.92	0.93		0.95
	$B_{\text{none,GR}}$	0.78	0.88	0.58	0.74		0.93
	$B_{\text{none,IR}}$	0.78	0.87	0.38	0.79		0.82
	$B_{\text{none,PR}}$	0.78	0.88	0.58	0.74		0.93
	$B_{\text{case,GR}}$	0.90	0.91	0.71	0.75		0.93
	$B_{\text{case,IR}}$	0.98	0.98	0.84	0.94		0.99
	$B_{\text{case,PR}}$	0.90	0.92	0.71	0.75		0.94
	$B_{\eta,\text{GR}}$	0.93	0.95	0.94	0.93		0.95
	$B_{\eta,\text{IR}}$	0.94	0.95	0.97	0.94		0.88
	$B_{P\eta,\text{PR}}$	0.94	0.95	0.94	0.94		0.95

Coverage rate from 90% to 100% is typeset in bold font.
CI, confidence interval; REML, restricted maximum likelihood.

residual-alone bootstraps had very poor coverage rates for all parameters (<0.6) except for the SE of σ , whereas the case bootstrap and the bootstraps of both random effects and residuals provided good coverage rates (close to the nominal value of 0.95). According to our predefined criteria, four methods showed good performance: $B_{\text{case,none}}$, $B_{\eta,\text{GR}}$, $B_{\eta,\text{IR}}$, and $B_{P\eta,\text{PR}}$. These methods remained the best methods for the sparse design and the large error design, with smaller relative bias for the parameter estimates and their SEs and better coverage rates. The simulation results also showed that the asymptotic approach provided good estimates for parameters and SEs as well as good coverage rates as do the bootstrap candidates. In this simple setting, the convergence rate was high: nearly all runs on the simulated datasets converged (respectively, 100%, 99.8%, and 99.9% for the rich, sparse, and large error designs). Convergence was also close to 100% when applied to the bootstrap samples in the rich (100%) and large

error (99.9%) design, whereas for the sparse design, convergence rates were slightly lower and more dependent on the bootstrap method, going from around 90.2% for the bootstraps combining case and residuals, to 96.7% for the case bootstrap or the bootstraps of both random effects and residuals.

In this study, we also evaluated the influence of transformation of random effects and residuals using the ratio between the estimated and empirical variance-covariance matrices. Nontransformed and transformed resampling were compared for all nonparametric bootstrap methods except for the $B_{\text{case,none}}$ where no transformation was needed. The results of this transformation in the rich design are presented in Figure 3. We found that these corrections improved significantly the estimate of σ , its SE as well as the coverage rate for all applied methods. However, they only improved the estimates and the coverage rates of other variance parameters for $B_{\eta,\text{GR}}$ and $B_{\eta,\text{IR}}$.

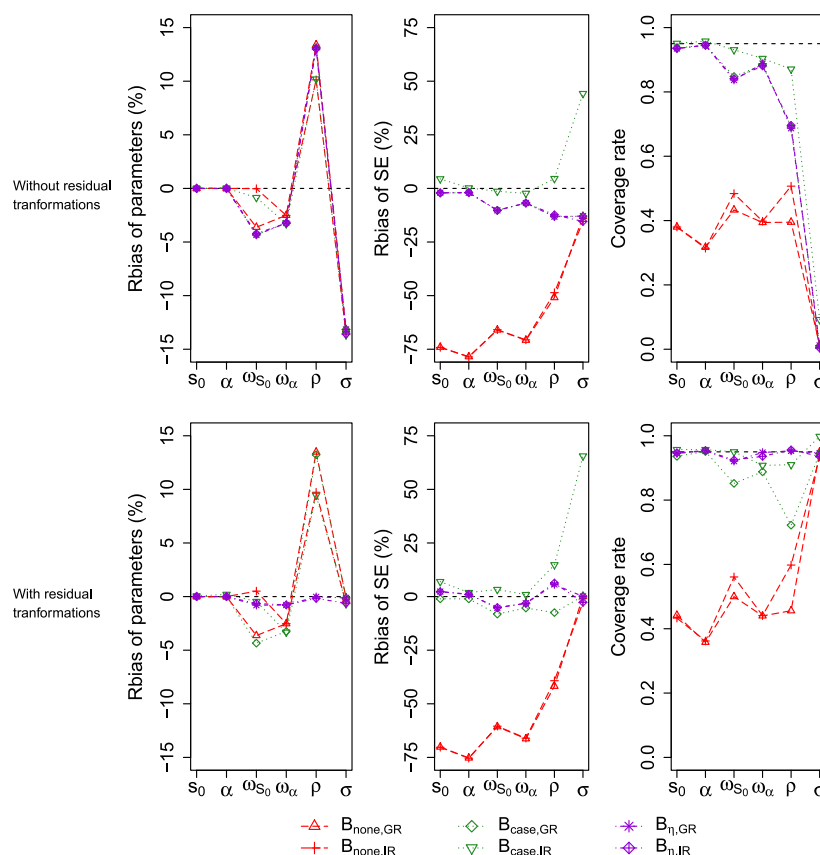


Figure 3. Relative bias of parameter estimates by restricted maximum likelihood (left), relative bias of standard errors (middle), and coverage rate of 95% confidence interval (right), for the nonparametric bootstrap methods without (top) and with (bottom) transformations of random effects and residuals in the rich design ($N = 100$, $n = 7$, $\sigma = 5.86$).

The results for the bootstrap candidates in the different evaluated designs, providing a more contrasted evaluation of their performance, were shown in Figure S1 in the Appendix. All the bootstrap candidates provided good parameter estimates in all evaluated designs with the relative bias within $\pm 5\%$, except for a higher bias on σ (-6.50%) observed for the $B_{\eta,IR}$ in the sparse design. The relative bias for the variance estimates were higher for the sparse and the large error designs compared with those of the rich design, whereas the relative bias of the fixed effects remained very small ($<0.1\%$). All the bootstrap candidates estimated correctly the SE of parameters, with relative biases ranging from -5% to 5% , observed in both rich and large error designs. They estimated less correctly the SE of variance parameters in the sparse design with relative biases ranging from -10% to 10% , especially for the case bootstrap. The boxplots of the SEs of all parameter estimates obtained by the bootstrap candidates in all evaluated designs are shown in Figure S2 in the Appendix. The range of bootstrap SEs across the $K = 1000$ replications did not show any practically relevant differences across the bootstrap methods, with the exception of SE of σ in the sparse design. A good coverage rate was obtained for all parameters in the rich design setting, as well as the design with large error; the one exception is a low coverage rate for σ observed for $B_{\eta,IR}$ in the design with large error. In the sparse design, the bootstrap candidates provided lower coverage rate for variance parameters. Across the different

designs, the $B_{P\eta,PR}$ worked slightly better than the $B_{case,none}$ and the $B_{\eta,GR}$ performed better than the $B_{\eta,IR}$.

To understand whether the estimation method influences the performance of the bootstrap methods, we compared the results of ML with those of REML. All the results of ML are given in the Appendix with Table S1 and Table S2. The difference between ML and REML in relative bias of parameter estimates and their SEs and coverage rate for all bootstrap methods and the asymptotic method in three evaluated designs are shown in Figure S3 in the Appendix. There was no difference in estimation of fixed effects and σ between two methods. However, the variance parameters (ω_{s_0} and ω_{α}) given by ML were slightly less well estimated compared with those of REML with increase of 0.5% to 2% in relative bias and had lower coverage rate in all evaluated designs. The difference between ML and REML in estimation of variance parameters was more apparent in the sparse design, resulting in bigger difference in their coverage rates (2% to 5%) observed for all bootstrap methods. In terms of SE estimation, ML and REML provided similar estimates, except for a small difference in relative bias ($<2\%$) in the sparse design. On the basis of our criteria about relative bias, REML was better than ML by providing unbiased estimates for almost parameters obtained by four bootstrap candidates. REML also improved the coverage rate of variance parameters in the sparse design while it provided similar estimates of SEs compared with ML.

Table IV. Parameter estimates by REML and their standard errors (SE) obtained by the asymptotic method, the bootstrap candidates, and the stratified bootstrap in the real dataset.

Method	Parameter estimates						SE					
	S_0	α	ω_{S_0}	ω_α	ρ	σ	S_0	α	ω_{S_0}	ω_α	ρ	σ
Asymptotic (ML)	23.99	13.97	11.08	12.80	0.63	5.86	1.12	1.40	0.84	1.44	0.08	0.17
Asymptotic (REML)	23.98	14.01	11.13	12.93	0.63	5.86	1.12	1.41	0.85	1.46	0.08	0.17
$B_{\text{case},\text{none}}$	24.02	14.10	11.10	12.99	0.63	5.81	1.11	1.82	0.78	2.04	0.10	0.54
$B_{\text{case},\text{none}}^{\text{strat}}$	24.01	14.11	11.07	13.03	0.63	5.84	1.02	1.58	0.77	1.80	0.10	0.54
$B_{\eta,\text{GR}}$	23.97	14.02	10.64	11.68	0.78	5.84	1.08	1.27	0.74	0.94	0.08	0.37
$B_{\eta,\text{IR}}$	24.07	14.01	10.70	11.87	0.74	5.77	1.10	1.33	0.78	1.10	0.11	0.35
$B_{P\eta,\text{PR}}$	23.97	14.01	11.07	12.83	0.63	5.86	1.13	1.42	0.82	1.10	0.09	0.16

REML, restricted maximum likelihood; ML, maximum likelihood.

4.2. Application to real data

The asymptotic parameter estimates by ML and their SEs for the real dataset were given in the first row of Table IV; these were taken as the true parameter values used in the simulation study reported earlier. The parameter estimates by REML and the SEs of all parameters obtained by the asymptotic method and the four bootstrap candidates for the real dataset are also shown in Table IV. They provided similar values for both fixed-effect and variance parameters that were also close to the asymptotic estimates. However, there were some differences for the estimation of SE of α , ω_α , and σ between the bootstrap methods. The $B_{\text{case},\text{none}}$ yield the highest SE for these parameters, whereas the $B_{P\eta,\text{PR}}$ and the asymptotic estimates were very similar, which were both different from the remaining bootstrap methods. This finding was not observed in the simulation study, where balanced designs with the same number of observations per subject were used and the residuals were normally distributed. The distribution of the empirical residuals in the real dataset was investigated and a nonnormal distribution was confirmed by the Shapiro–Wilk normality test. The similar results of ML for the real data are also presented in Table S3 in the Appendix.

To investigate the effect of the unbalance in the design, we stratified the case bootstrap based on the number of observations per subject (nobs). The real dataset was divided into three groups before bootstrapping: group 1 ($N = 40$ subjects with nobs ≤ 5), group 2 ($N = 38$ subjects with nobs > 5 and ≤ 10), and group 3 ($N = 31$ subjects with nobs > 10). The bootstrap samples were then built by sampling subjects within each group, keeping the same number of subjects from each group. The results obtained from this analysis are also given in Table IV. The case bootstrap with stratification $B_{\text{case},\text{none}}^{\text{strat}}$ gave similar values for parameters as the case bootstrap and other bootstrap methods. In terms of SE of parameters, this method provided smaller values for SEs of α and ω_α and reduced the difference on SE estimation of these parameters with other bootstrap methods, which may indicate that the case bootstrap is more sensitive to unbalanced designs; on the other hand, stratification hardly affected the SE of σ .

5. DISCUSSION

In this paper, we evaluated different bootstrap approaches for estimating SEs and CIs of parameters in LMEMs with homoscedastic error. The proposed bootstraps take into account two levels of variability in the longitudinal data: BSV and RUV. They were also

compared with the residual bootstrap, which resamples only one level of variability, and to the case bootstrap where the whole vector of observations is resampled.

Our simulations showed that bootstrapping only residuals underestimated greatly the SEs of parameters, except for σ , and provided poor coverage rates. This finding is to be expected because the large BSV for two parameters in the evaluated designs were not taken into account. On the contrary, the case bootstrap performed well as it provided nonbiased parameter estimates and SEs as well as good coverage rates for all parameters. Moreover, according to Van der Leeden *et al.*, it makes sense to resample only the individuals and collect the related observations from the original dataset when bootstrapping cases for the repeated measures data [14, 26]. Our results support the implication that the RUV is somewhat taken into account in this method; resampling cases preserves both the BSV and the RUV. It is in agreement with the worse performance of combining residual bootstrap with case bootstrap, in which the RUV is considered already resampled.

Another important result of this study is the good performance of bootstrapping both random effects and residuals, either in a nonparametric or in a parametric way. They worked as well as the case bootstrap. The incorporation of random effects into the classical residual bootstrap plays therefore a very important role for bootstrapping in mixed-effects models context, especially when the BSV is much higher than the RUV. This approach was proposed in various studies, such as for the resampling of multi-level data [14, 23], times series data [10], and more recently for longitudinal data [15]. It can be an alternative method to the case bootstrap when the model is correct (for nonparametric version) or both model and assumptions on distributions of parameters are correct (for parametric version). To our knowledge, there has not been any simulation study in the literature that compares this approach to the most commonly used method, the case bootstrap for mixed-effects models with longitudinal data.

Bootstrapping the raw random effects and residuals does not take into account variance underestimation, leading to shrinkage in the individual parameter estimates. To account for this issue, we employed the correction using the ratio between estimated and empirical variance–covariance matrix for the random effects and the residuals. It was shown to be an appropriate method for LMEMs because of the improvement of estimation for variance components. These ratios account for the degree of two shrinkages: η -shrinkage and ϵ -shrinkage, which quantify the amount of information in the individual data about the parameters [27–29].

When the data is not informative, the random effects and residuals are shrunk toward 0, and high degree of η -shrinkage and ϵ -shrinkage will be obtained. Sampling in the raw distribution will therefore underestimate the actual level of variability in the data, whereas correcting both empirical random effects and residuals for shrinkage restores this level. This idea of accounting for the difference between the estimated and empirical variance of residuals through an estimate of the shrinkage was proposed in bootstrapping ordinary linear models [21] and was extended for the two levels of variability found in mixed models by Wang *et al.* [23].

The performance of different bootstrap methods was evaluated under different conditions: the rich design in which all parameters can be well estimated, the sparse design in which the variance parameters are less well estimated, and the large error design in which the RUV is as important as the BSV to see whether the residual bootstraps work better. The convergence was obtained for almost all bootstrap datasets (90% to 100%), which should not provide substantial bias for estimating the uncertainty of parameters. The case bootstrap and three bootstrap methods where both random effects and residuals were resampled remained the best methods and selected as bootstrap candidates for LMEMs. The purpose of this work was not to determine which was the best method overall, but to eliminate bootstrap methods that do not perform well even with LMEMs. We did note that the global residual bootstrap was slightly better than individual residual bootstrap in the sparse and large error designs, especially in estimating σ , which is consistent with the noncorrelated structure of residuals. In addition, the distribution of resampled residuals obtained by the global residual bootstrap was slightly closer to the original distribution of residuals. The parametric bootstrap performed best across three evaluated designs, but it requires the strongest assumptions (good prior knowledge about model structure and distributions of parameters). If the model is misspecified and the assumptions of normality of random effects and residuals are not met, this method may not be robust. In practice, one of the main reasons for using bootstrap is the uncertainty of distribution assumption, the nonparametric bootstrap may therefore preferable to the parametric bootstrap in most applications [9].

We also investigated using ML instead of REML when performing the estimation step. The difference in performance was small with less than 5%, but there was slightly more bias with ML especially for the estimation of the variance parameters, which was most apparent in the sparse design where variances were less well estimated. This means that we may expect this to be true also for nonlinear mixed effect models where ML is more often used than REML for parameter estimation, although this needs to be verified in the nonlinear setting. This finding should be also explored further in other settings, because the superiority of REML over ML becomes more apparent as the number of fixed effect increases, especially when the number of subjects is limited [3].

The number of bootstrap replicates (B) depends on the estimation that we want to obtain. In simple regression models, B is recommended to be at least 100 for SE estimation and at least 1000 in the case of CI estimation [30, 31]. In this simulation study, 1000 bootstrap replicates were thought to be large enough to obtain both bootstrap SE and bootstrap CI for all bootstrap methods with LMEMs. Note that, to estimate directly the quantiles for 95% CI without interpolation, $B = 999$ should be used instead of 1000 in the future work [7]. In addition, further evaluation on choosing the number of bootstrap will be studied, especially

when bootstrapping on nonLMEMs is much time-consuming and less stable.

Although in the simulation studies the performance of the case bootstrap was similar to that of the bootstrap methods resampling both random effects and residuals, when these methods were applied to the real dataset, the case bootstrap estimated a much larger SE for both α and its variability, and there was also smaller differences in the estimates of σ between the different bootstraps. We found a good agreement between the $B_{P\eta,PR}$ and the asymptotic method, which were both different from the remaining methods. The difference between $B_{P\eta,PR}$ and $B_{\eta,GR}$ could come from different assumptions on distributions, because for the former, we sample in a normal distribution, whereas the empirical residuals used for resampling in $B_{\eta,GR}$ were in fact not normally distributed. The difference between $B_{case,none}$ and $B_{\eta,GR}$ or $B_{\eta,IR}$ might be due to the unbalanced design of the real dataset in which patients have different number of observations, a structure which is preserved by the residual bootstraps but not by the case bootstrap. Unbalanced designs, therefore, may be more challenging for the case bootstrap than for other bootstrap methods [7, 32], and stratification has been proposed to handle such situations; for example, when a study includes rich and sparse data, it is recommended to resample from both groups to maintain a similar structure in the bootstrap samples [33]. In this study, we tried to apply the stratified case bootstrap to the real dataset. The stratification could explain a part of the difference between the case bootstrap and other bootstrap candidates by decreasing the SEs of α and ω_α to the values closer to those obtained by other methods.

In conclusion, in this study, we found that the case bootstrap performs as well as the nonparametric/parametric bootstrap of random effects and residuals in LMEMs with balanced designs and homoscedastic error. However, the residual bootstraps always generate datasets with the same design as the original data and would be expected to perform better in situations where the design is not similar for every individual. This could be the explanation for the discrepancy between the bootstraps seen in the application to the real data from the Parkinson study. We now plan to compare these methods for nonLMEMs, addressing the issues of heteroscedasticity, the nonlinearity of model, and exploring the influence of designs with stratified bootstrap.

Acknowledgements

The authors would like to thank the Parkinson Study Group for providing us the data from Parkinson's study as well as IFR02 and Hervé Le Nagard for the use of the 'centre de biomodélisation'. We also thank the reviewer for a careful review and insightful comments that helped us to improve this paper.

During this work, Hoai-Thu Thai was supported by a research grant from Drug Disposition Department, Sanofi, Paris.

REFERENCES

- [1] Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. John Wiley & Sons: New York, 2004.
- [2] Pinheiro JC, Bates DM. *Mixed-effects models for s and s-plus*. Springer: New York, 2000.
- [3] Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. Springer: New York, 2000.
- [4] Efron B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 1979; **7**(1):1–26.

- [5] Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Chapman & Hall: New York, 1994.
- [6] Shao J, Tu D. *The jackknife and bootstrap*. Springer: New York, 1995.
- [7] Davison AC, Hinkley DV. *Bootstrap methods and their application*. Cambridge University Press: Cambridge, 1997.
- [8] MacKinnon JB. Bootstrap methods in econometrics. *The Economic Record* 2006; **82**(s1):S2–S18.
- [9] Wehrens R, Putter H, Buydens Lutgarde M C. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems* 2000; **54**(1):35–52.
- [10] Ocana J, El Halimi R, Ruiz de Villa MC, Sanchez JA. Bootstrapping repeated measures data in a nonlinear mixed-models context. *Mathematics Preprint Series* 2005; **367**:1–31.
- [11] Flachaire E. Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics and Data Analysis* 2005; **49**(2):361–376.
- [12] Halimi R. *Nonlinear mixed-effects models and bootstrap resampling: Analysis of non-normal repeated measures in biostatistical practice*. VDM Verlag Dr. Möjller: Berlin, 2005.
- [13] Das S, Krishen A. Some bootstrap methods in nonlinear mixed-effect models. *Journal of Statistical Planning and Inference* 1999; **75**(2):237–245.
- [14] Van der Leeden R, Busing FMTA, Meijer E. Bootstrap methods for two-level models. *Technical Report PRM 97-04*, Leiden University, Department of Psychology, Leiden, 1997.
- [15] Wu H, Zhang J-T. The study of long-term hiv dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine* 2002; **21**(23):3655–3675.
- [16] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**(4):963–974.
- [17] Ette EI. Stability and performance of a population pharmacokinetic model. *Journal of Clinical Pharmacology* 1997; **37**(6):486–495.
- [18] Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* 1986; **14**(4):1261–1295.
- [19] Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine* 2000; **19**(9):1141–1164.
- [20] Carpenter JR, Goldstein H, Rasbash J. A novel bootstrap procedure for assessing the relationship between class size and achievement. *Applied Statistics* 2003; **52**:431–443.
- [21] Moulton LH, Zeger SL. Bootstrapping generalized linear models. *Computational Statistics and Data Analysis* 1991; **11**(1):53–63.
- [22] Efron B. *The jackknife, the bootstrap and other resampling plans*. Society of Industrial and Applied Mathematics: Philadelphia, 1982.
- [23] Wang J, Carpenter JR, Kepler MA. Using SAS to conduct nonparametric residual bootstrap multilevel modeling with a small number of groups. *Computer Methods and Programs in Biomedicine* 2006; **82**(2):130–143.
- [24] Holford NHG, Chan PLS, Nutt JG, Kiebertz K, Shoulson I, Parkinson Study Group. Disease progression and pharmacodynamics in Parkinson disease—evidence for functional protection with levodopa and other treatments. *Journal of Pharmacokinetics and Pharmacodynamics* 2006; **33**(3):281–311.
- [25] Bates DM, Pinheiro JC. Computational methods for multilevel models. *Technical Memorandum BL0112140-980226-01TM*, Bell Labs, Lucent Technologies, Murray Hill, NJ, 1998.
- [26] De Leeuw J, Meijer E. *Handbook of multilevel analysis*, chap. Resampling multilevel models. Springer: New York, 2007.
- [27] Ette EI, Williams PJ. *Pharmacometrics: the science of quantitative pharmacology*. Wiley-Blackwell: New Jersey, 2007.
- [28] Karlsson MO, Savic RM. Diagnosing model diagnostics. *Clinical Pharmacology and Therapeutics* 2007; **82**(1):17–20.
- [29] Savic RM, Karlsson MO. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *The AAPS Journal* 2009; **11**(3):558–569.
- [30] Chernick MR. *Bootstrap methods: A guide for practitioners and researchers*, (2nd edn). John Wiley & Sons: New Jersey, 2007.
- [31] Bonate PL. *Pharmacokinetic-pharmacodynamic modeling and simulation*, (2nd edn). Springer: New-York, 2011.
- [32] Davison AC, Kuonen D. An introduction to the bootstrap with applications in R. *Statistical Computing and Statistical Graphics Newsletter* 2002; **13**(1):6–11.
- [33] Lindbom L, Pihlgren P, Jonsson EN. PsN-Toolkit—A collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Computer Methods and Programs in Biomedicine* 2005; **79**(3):241–257.

APPENDIX

Supplementary tables

Table S1: Relative bias of parameter estimates by ML and their standard errors (SE) for the asymptotic method and the bootstrap methods in the 3 studied designs

Design	Method	Relative bias of parameters (%)					Relative bias of SE (%)				
		S ₀	α	ω_{S_0}	ω_{α}	ρ	S ₀	α	ω_{S_0}	ω_{α}	ρ
Rich	Asymptotic	0.05	0.19	-0.68	-0.57	-0.09	2.28	1.11	-3.20	-1.51	4.46
	B _{case} , none	0.00	0.01	-0.85	-0.81	0.15	2.31	1.17	-5.89	-3.42	5.07
	B _{none} , GR	0.00	0.00	-3.71	-2.58	13.77	-70.16	75.26	-60.65	-66.29	-41.76
	B _{none} , IR	0.00	0.00	0.47	-2.58	9.97	-70.29	75.37	-60.43	-66.05	-39.15
	B _{none} , PR	0.00	0.00	-3.71	-2.57	13.76	-70.15	75.22	-60.63	-66.22	-41.73
	B _{case} , GR	0.00	0.01	-4.42	-3.31	13.52	-1.11	-1.12	-8.22	-5.37	-7.76
	B _{case} , IR	0.01	0.02	-0.36	-3.37	10.15	6.94	1.82	2.80	0.41	14.70
	B _{case} , PR	0.00	-0.01	-4.43	-3.31	13.49	-1.25	-1.32	-8.00	-5.36	-7.61
	B _{η} , GR	0.00	-0.01	-1.35	-1.31	0.16	1.83	0.60	-5.73	-3.56	5.68
	B _{η} , IR	0.00	0.00	-1.19	-1.30	0.01	1.75	0.49	-5.53	-3.59	6.39
Sparse	B _{Pη} , PR	0.00	0.00	-0.86	-0.82	0.15	2.24	1.06	-3.17	-1.57	6.02
	Asymptotic	0.08	0.48	-2.85	-2.61	1.48	1.23	-1.64	-2.94	-4.14	-5.24
	B _{case} , none	-0.01	-0.02	-2.63	-2.62	0.66	1.05	-1.89	-9.08	-9.81	-7.34
	B _{none} , GR	0.00	0.00	-5.36	-3.28	18.81	-62.92	-72.24	-51.06	-62.03	-41.01
	B _{none} , IR	0.00	0.00	6.82	-3.41	9.87	-64.35	-73.44	-50.72	-63.21	-32.67
	B _{none} , PR	0.00	-0.01	-5.36	-3.28	18.79	-62.73	-72.11	-50.61	-62.01	-40.60
	B _{case} , GR	0.02	0.03	-7.60	-5.61	17.35	-3.45	-4.62	-11.28	-11.20	-17.67
	B _{case} , IR	0.01	0.02	4.46	-5.70	8.60	12.26	-1.37	6.40	-5.37	9.28
	B _{case} , PR	-0.01	0.02	-7.54	-5.61	17.33	-3.41	-4.52	-11.06	-11.06	-17.21
	B _{η} , GR	0.02	0.02	-4.26	-4.23	0.79	-0.50	-3.27	-8.29	-9.90	-4.57
Large error	B _{η} , IR	-0.01	0.01	-2.65	-4.20	-0.94	-0.90	-3.62	-6.78	-10.39	0.35
	B _{Pη} , PR	0.00	0.03	-2.71	-2.67	0.59	1.25	-1.57	-3.39	-4.43	-4.28
	Asymptotic	0.02	-0.08	-1.54	-2.31		3.26	-2.37	-2.69	-4.09	1.06
	B _{case} , none	0.00	-0.01	-1.20	-3.64		-3.42	-2.49	-3.52	1.00	-0.50
	B _{none} , GR	-0.01	0.01	-10.80	-25.67		-37.52	-20.83	-29.91	6.12	-1.60
	B _{none} , IR	-0.01	0.01	18.79	-25.98		-38.60	-22.33	-35.03	16.20	-6.15
	B _{none} , PR	0.00	0.00	-10.81	-25.66		-37.48	-20.86	-29.83	6.21	-1.21
	B _{case} , GR	-0.01	-0.01	-11.53	-26.31		-12.28	-12.58	-7.37	12.05	-1.43
	B _{case} , IR	0.01	0.00	17.58	-29.31		19.46	15.80	19.76	61.24	60.36
	B _{case} , PR	0.00	0.00	-11.52	-26.35		-12.27	-12.43	-7.27	12.14	-1.26
Relative bias within $\pm 5\%$ is typeset in bold font	B _{η} , GR	0.00	-0.01	-1.70	-4.22		-3.67	-2.45	-3.20	1.97	0.73
	B _{η} , IR	0.01	0.00	2.68	-4.28		-4.34	-3.85	-2.14	3.63	-4.23
	B _{Pη} , PR	0.00	0.01	-1.22	-3.72		-3.28	-2.37	-2.08	2.41	0.90

Table S2: Coverage rate of the 95% CI of parameters estimates by ML obtained by the asymptotic method and the bootstrap methods in the 3 studied designs

Design	Method	Coverage rate of the 95% CI of parameters					
		S_0	α	ω_{S_0}	ω_α	ρ	σ
Rich	Asymptotic	0.95	0.96	0.93	0.94	0.95	0.95
	B _{case,none}	0.94	0.96	0.91	0.93	0.95	0.94
	B _{none,GR}	0.44	0.36	0.50	0.44	0.44	0.95
	B _{none,IR}	0.43	0.36	0.56	0.44	0.58	0.93
	B _{none,PR}	0.43	0.36	0.49	0.44	0.44	0.94
	B _{case,GR}	0.94	0.95	0.84	0.88	0.71	0.94
	B _{case,IR}	0.96	0.96	0.95	0.90	0.90	1.00
	B _{case,PR}	0.94	0.95	0.84	0.88	0.70	0.94
	B $_{\eta,GR}$	0.94	0.95	0.91	0.93	0.95	0.95
	B $_{\eta,IR}$	0.94	0.95	0.92	0.93	0.96	0.94
	B $P_{\eta,PR}$	0.95	0.96	0.93	0.94	0.96	0.94
Sparse	Asymptotic	0.94	0.94	0.92	0.91	0.90	0.94
	B _{case,none}	0.94	0.93	0.87	0.87	0.92	0.90
	B _{none,GR}	0.53	0.39	0.57	0.49	0.63	0.95
	B _{none,IR}	0.51	0.38	0.67	0.46	0.83	0.92
	B _{none,PR}	0.53	0.40	0.58	0.49	0.63	0.93
	B _{case,GR}	0.93	0.93	0.80	0.83	0.82	0.94
	B _{case,IR}	0.96	0.94	0.98	0.86	0.99	0.99
	B _{case,PR}	0.93	0.93	0.80	0.82	0.83	0.93
	B $_{\eta,GR}$	0.94	0.93	0.87	0.86	0.94	0.95
	B $_{\eta,IR}$	0.94	0.94	0.91	0.86	0.96	0.91
	B $P_{\eta,PR}$	0.94	0.94	0.90	0.89	0.94	0.94
Large error	Asymptotic	0.93	0.95	0.94	0.98		0.95
	B _{case,none}	0.93	0.94	0.92	0.93		0.95
	B _{none,GR}	0.78	0.88	0.56	0.70		0.93
	B _{none,IR}	0.78	0.87	0.39	0.77		0.81
	B _{none,PR}	0.78	0.88	0.56	0.70		0.93
	B _{case,GR}	0.90	0.92	0.68	0.72		0.93
	B _{case,IR}	0.98	0.98	0.84	0.94		0.99
	B _{case,PR}	0.90	0.92	0.69	0.72		0.94
	B $_{\eta,GR}$	0.93	0.95	0.92	0.93		0.95
	B $_{\eta,IR}$	0.93	0.94	0.97	0.93		0.88
	B $P_{\eta,PR}$	0.94	0.94	0.93	0.93		0.95

Coverage rate from 90% to 100% is typeset in bold font

Table S3: Parameter estimates by ML and their standard errors (SE) obtained by the asymptotic method, the bootstrap candidates and the stratified bootstrap in the real dataset

Method	Parameter estimates						SE					
	S_0	α	ω_{S_0}	ω_α	ρ	σ	S_0	α	ω_{S_0}	ω_α	ρ	σ
Asymptotic	23.99	13.97	11.08	12.80	0.63	5.86	1.12	1.40	0.84	1.44	0.08	0.17
$B_{\text{case,none}}$	24.01	13.99	10.98	12.75	0.63	5.85	1.12	1.84	0.83	2.00	0.10	0.53
$B_{\text{case,none}}^{\text{strat}}$	23.94	13.96	10.97	12.83	0.63	5.83	1.02	1.56	0.78	1.79	0.10	0.52
$B_{\eta,\text{GR}}$	23.98	14.02	10.56	11.52	0.79	5.84	1.08	1.26	0.76	0.90	0.08	0.35
$B_{\eta,\text{IR}}$	23.98	13.97	10.61	11.74	0.75	5.78	1.11	1.31	0.75	1.06	0.11	0.36
$B_{P\eta,\text{PR}}$	23.98	13.97	11.03	12.70	0.64	5.86	1.12	1.36	0.83	1.13	0.08	0.17

Supplementary figures

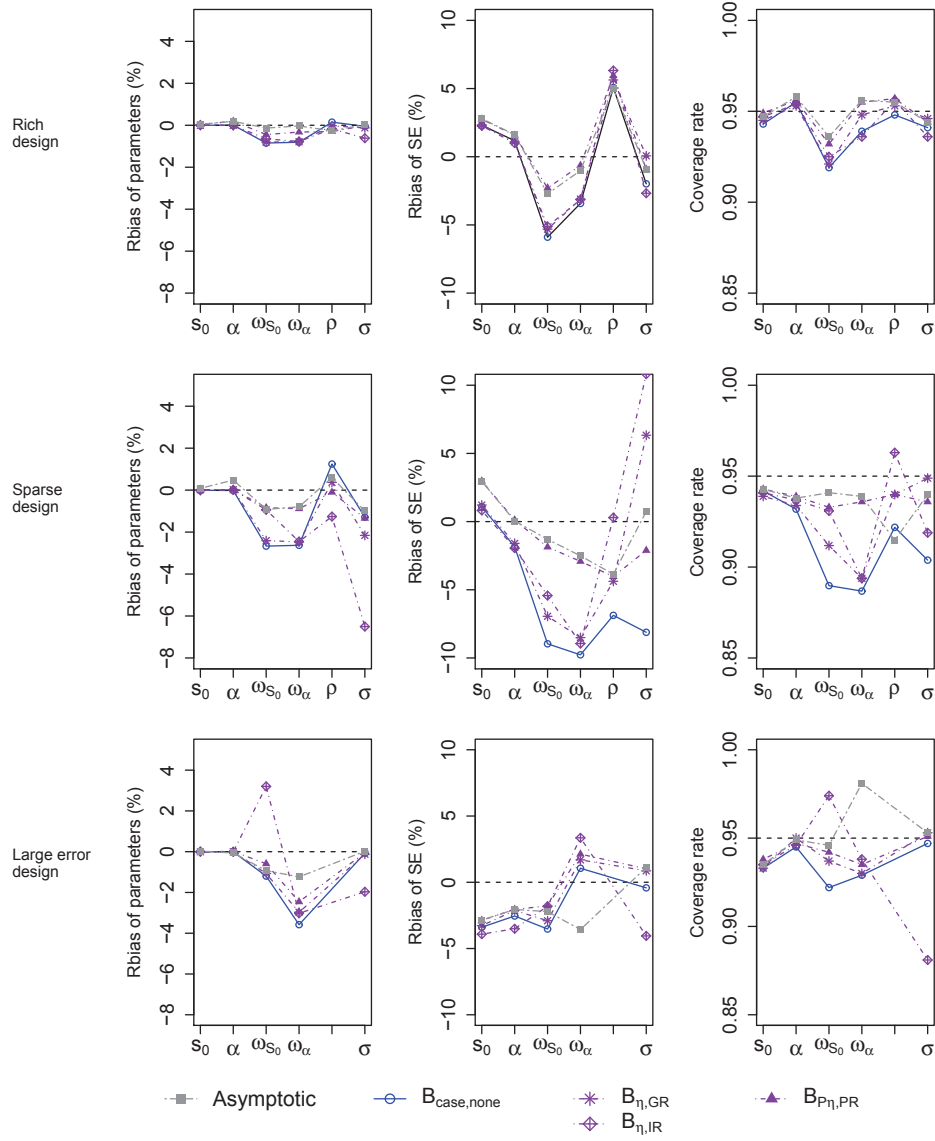


Figure S1: Relative bias of parameter estimates by REML (left), relative bias of standard errors (middle) and coverage rate of 95% CI (right), for the asymptotic method and the bootstrap candidates in the rich design ($N=100$, $n=7$, $\sigma=5.86$) (top), the sparse design ($N=30$, $n=3$, $\sigma=5.86$) (middle) and the large error design ($N=100$, $n=7$, $\sigma=17.5$, $\rho=0$) (bottom). The same scales were plotted to compare bootstrap methods with each other in the 3 studied designs.

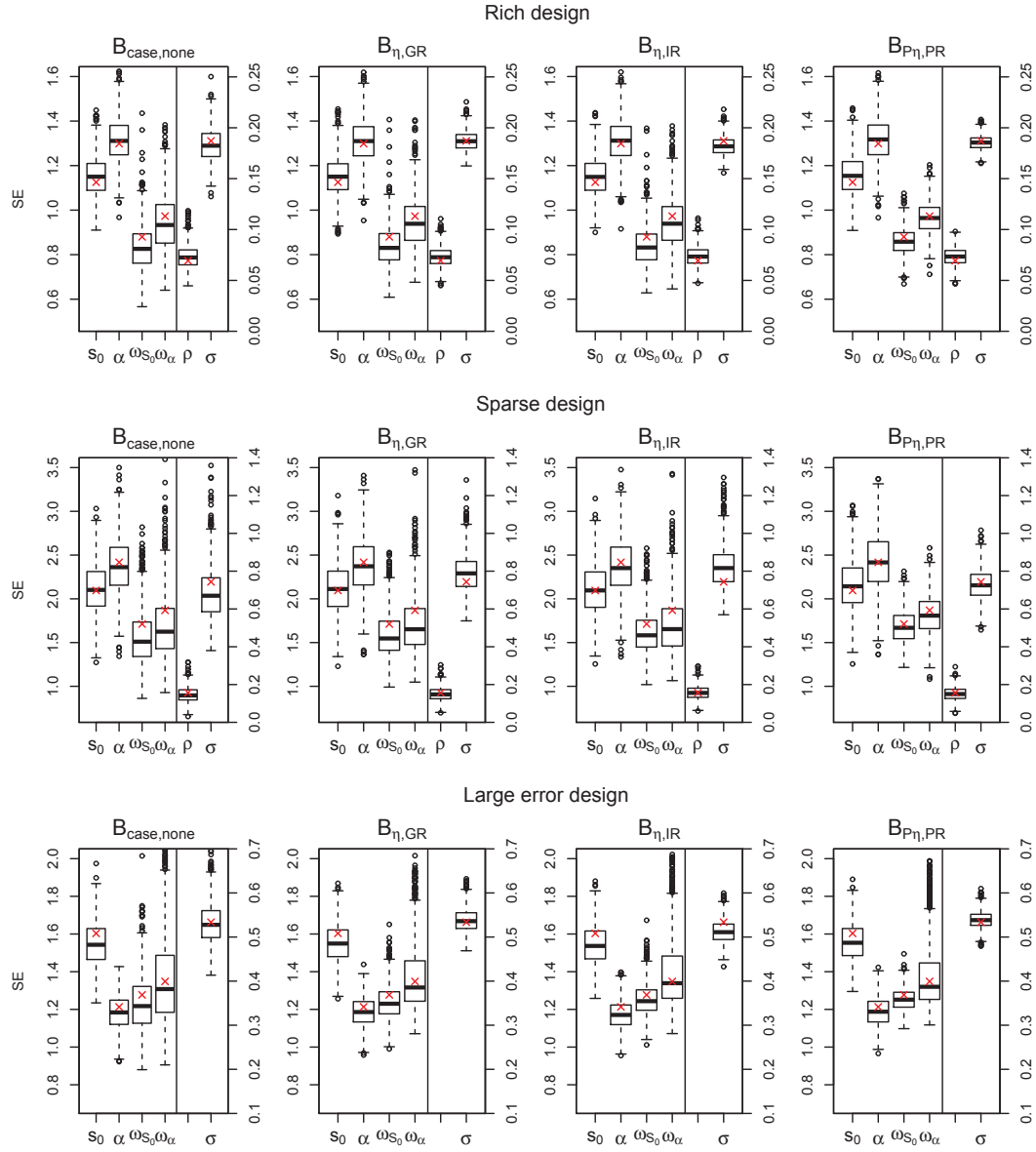


Figure S2: Boxplot of standard errors (SE) of parameters by REML obtained by the bootstrap candidates in the 3 studied designs: the rich design ($N=100$, $n=7$, $\sigma=5.86$) (top), the sparse design ($N=30$, $n=3$, $\sigma=5.86$) (middle) and the large error design ($N=100$, $n=7$, $\sigma=17.5$, $\rho=0$) (bottom). A second y-axis with smaller scale was plotted for ρ and σ in the right side of each boxplot. The empirical values obtained by K=1000 simulations were presented as red crosses.

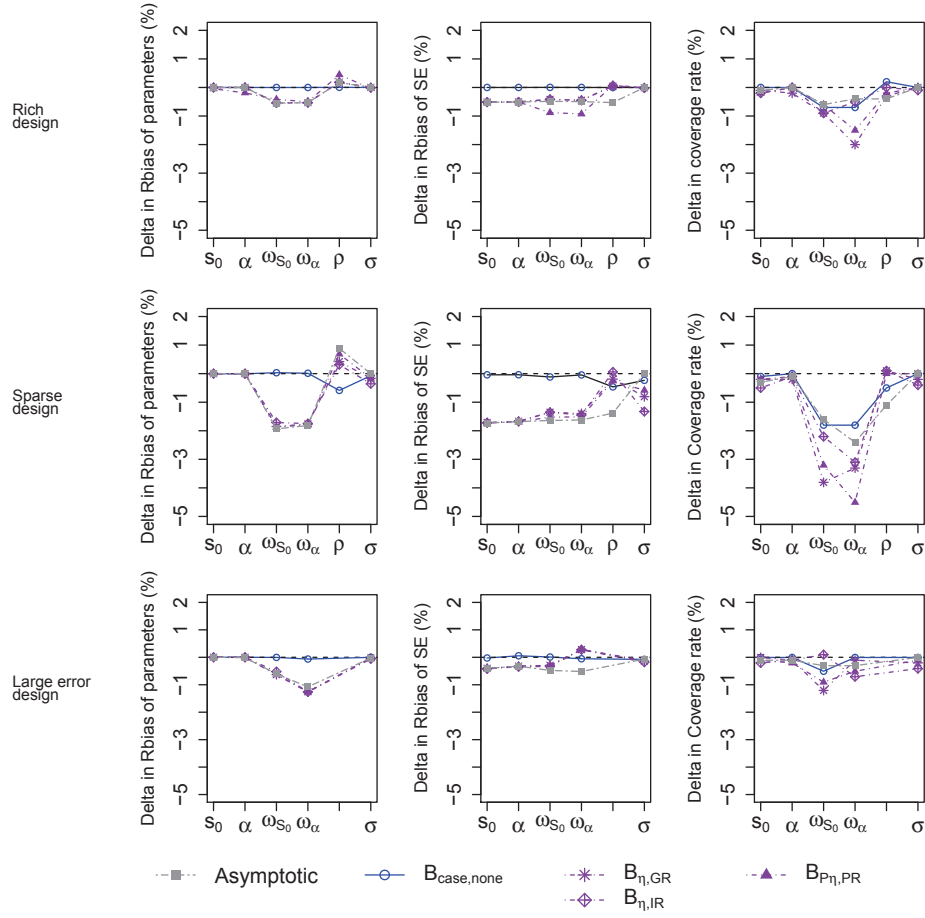


Figure S3: Difference (delta) between ML and REML (ML-REML) in relative bias of parameter estimates (left), relative bias of standard errors (middle) and coverage rate of 95% CI (right), for the asymptotic method and the bootstrap methods in the rich design (N=100, n=7, $\sigma=5.86$) (top), the sparse design (N=30, n=3, $\sigma=5.86$) (middle) and the large error design (N=100, n=7, $\sigma=17.5$, $\rho=0$) (bottom).

3.2 Evaluation des approches bootstrap dans les modèles non linéaires à effets mixtes

3.2.1 Résumé

Dans le travail précédent concernant les approches bootstrap dans les MLEM, les meilleures performances ont été trouvées pour le bootstrap par paires et les bootstraps non paramétrique/paramétrique des effets aléatoires et des résidus au niveau de l'ensemble des sujets (THAI et al., 2013).

Dans ce travail, nous les avons évalués dans le cadre des modèles non linéaires à effets mixtes (MNLEM) en présence de deux complexités supplémentaires (non-linéarité et hétéroscédasticité) par une étude de simulation. Cette étude a été basée sur deux essais cliniques de l'affibercept chez les patients atteints de cancer (FREYER et al., 2012; RAMLAU et al., 2012). Après administration en perfusion, la PK de l'affibercept libre était décrite par un modèle à deux compartiments avec une élimination d'ordre 1. Pour les bootstraps non paramétrique/paramétrique des résidus, les résidus ont été normalisés à la même variance et puis rééchantillonnés comme dans le cas d'homoscédasticité (BONATE, 2011). Les performances de ces méthodes de bootstrap ont été évaluées dans deux protocoles équilibrés : un protocole d'échantillonnage fréquent ($N = 30$ patients/ $n = 9$ observations) et un protocole d'échantillonnage allégé ($N = 70/n = 4$). Nous avons également introduit une plus grande non-linéarité du modèle dans le protocole d'échantillonnage fréquent en utilisant une élimination Michaelis-Menten au lieu d'une élimination d'ordre 1. Les performances de ces méthodes de bootstrap ont également été évaluées dans un protocole déséquilibré comprenant les données fréquentes et éparées ($N_1=15/n_1=9$ et $N_2=75/n_2=2$) comme dans le jeu de données réelles. L'algorithme SAEM mis en œuvre dans MONOLIX 4.1.2 a été utilisé pour obtenir les estimations ML des paramètres du modèle. Les méthodes de bootstrap ont été étudiées par 100 réplifications pour chaque protocole et 999 échantillons bootstrap par réplification pour chaque méthode avec les mêmes critères d'évaluation dans les MLEM.

Nos simulations ont montré que le bootstrap par paires et le bootstrap non paramétrique/paramétrique des effets aléatoires et des résidus fonctionnent bien dans les protocoles équilibrés. Par rapport à l'approche asymptotique, elles fournissent une meilleure description de l'incertitude de certains paramètres, notamment ceux du modèle le plus non linéaire dans les protocoles d'échantillonnage fréquents. Cependant, elles fonctionnent moins bien dans le protocole allégé, en fournissant un biais très important pour la SE d'un paramètre estimé ayant une distribution asymétrique. Dans le protocole déséquilibré, le bootstrap par paires a également surestimé la SE de ce paramètre et le bootstrap non paramétrique des effets aléatoires et des résidus a surestimé la SE des paramètres de variance malgré la stratification basée sur les données fréquentes/éparées.

En revanche, la méthode asymptotique et le bootstrap paramétrique ont montré de bonnes estimations du SE. Ces résultats illustrent que les méthodes de bootstrap non paramétrique peuvent rencontrer des problèmes pratiques qui entraînent de mauvaises performances telles que les distributions asymétriques dans l'estimation des paramètres et les protocoles très déséquilibrés.

L'ensemble de ce travail fait également l'objet d'un article en cours de préparation pour la revue *Journal of Pharmacokinetics and Pharmacodynamics*.

3.2.2 Article 4 (soumis)

Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics

Hoai-Thu Thai · France Mentré · Nicholas H.G. Holford ·
Christine Veyrat-Follet · Emmanuelle Comets

Received: date / Accepted: date

Abstract Bootstrap methods are used to estimate the uncertainty of parameters in multi-level or linear mixed-effects models. Residual-based bootstrap methods which resample both random effects and residuals are an alternative approach to case bootstrap, which resamples the individuals. Most PKPD applications use the case bootstrap, for which software is available. In this study, we evaluated the performance of different bootstrap methods by a simulation study and compared them to that of an asymptotic method in estimating uncertainty of parameters in nonlinear mixed-effects models (NLMEM) with heteroscedastic error. This simulation was based on a real PK data collected from two clinical trials of aflibercept, an anti-angiogenic drug, in cancer patients. As expected, we found that the bootstrap methods provided better estimates of uncertainty for parameters in NLMEM with high nonlinearity compared to the asymptotic method. Overall, the case bootstrap proved the most robust method, and it was also fast and simple as it makes no assumptions on the model and preserves both between subject and residual variability in one resampling step. The nonparametric residual bootstrap also performed well, with a correction for variance underestimation before resampling. However, there are some situations that make the bootstraps perform worse than the asymptotic method, for example in unbalanced designs where the simple stratification seemed to be insufficient.

Keywords Bootstrap · Nonlinear mixed-effects models · Pharmacokinetics · Uncertainty of parameters · MONOLIX

1 Introduction

Nonlinear mixed-effects models (NLMEM), also called population analysis, have been widely used in the field of pharmacokinetics (PK) and pharmacodynamics (PD) to characterize the profile of drug concentrations or treatment response over time for a population. This approach not only provide the estimates of fixed-effect parameters in the studied population but also describe their variability quantified by the variance of the random effects. Population analysis was introduced by Lewis Sheiner and Stuart Beal and has now become an integral part of drug development since the first application of NLMEM in population PKPD in the late 1970s [1]. The parameters of NLMEM are often estimated by the maximum likelihood (ML) method.

H.T Thai · E. Comets · F.Mentré
INSERM, UMR 738, F-75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, UMR 738, F-75018 Paris, France
E-mail: hoai-thu.thai@inserm.fr

N.H.G Holford
Department of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand

H.T Thai · C.Veyrat-Follet
Drug Disposition Department, Sanofi, Paris, France

These models are complex, not only structurally but also statistically, with a number of assumptions about the model structure and variability distributions. The uncertainty of parameters in NLMEM is usually quantified by the standard errors (SE) obtained asymptotically by the inverse of the Fisher information matrix (\mathbf{M}_F) and by the asymptotic confidence intervals (CI) which are assumed to be normal and symmetric. However, this uncertainty might be biased when the assumption of asymptotic normality for parameter estimates and their SE is incorrect, for example when the sample-size is small or the model is more than trivially nonlinear. Sometimes, they cannot be even obtained due to identifiability issues in the model or numerical problems when evaluating the inverse of \mathbf{M}_F .

The bootstrap is an alternative method to assess the uncertainty of parameters without making strong distributional assumptions. It was first introduced by Efron (1979) for independent and identically distributed (iid) observations. The principal idea behind the bootstrap is to repeatedly resample the observed data with replacement to create new datasets having the same size as the original dataset, then fit each bootstrap dataset to construct the distribution of an estimator or a statistic of interest [2, 3]. In standard linear regression, the most simple and intuitive method is the case bootstrap which consists of resampling the pairs of observations with replacement. However, other bootstrap methods exist; for example the residual bootstrap and the parametric bootstrap [4, 5, 6, 7]. The residual bootstrap resamples the observed residuals obtained after model fitting then constructs the bootstrap samples. The parametric bootstrap adopts the principle of the residual bootstrap but simulates the residuals from the estimated distribution obtained by the fitting of the original data, e.g the normal distribution. In the mixed-effects models setting, the case bootstrap consists of resampling the whole vector of observations in one subject with replacement. The classical bootstrap methods used in linear regression with just one level of variability (residual) need be modified to take into account the characteristics of mixed-effects models with two levels of variability (between-subject and residual variability) [8]. Resampling random effects may be coupled with resampling residuals [8, 9, 10, 11]. The case bootstrap can be combined with the residual bootstrap [6].

In a previous study [12], we conducted a simulation study to evaluate different bootstrap methods that could be used for linear mixed-effects models (LMEM) with homoscedastic residual error, a simple case before moving to NLMEM. Resampling two levels of variability (case bootstrap coupled with residual bootstrap and random effect bootstrap coupled with residual bootstrap) was compared to resampling one level of variability (case alone or residual alone bootstrap). The nonparametric and parametric versions of these bootstrap methods were evaluated. Three balanced designs (rich, sparse and large error) were used to evaluate the respective importance of two levels of variability. The study demonstrated the adequate performance of the nonparametric/parametric random effect and residual bootstrap which resamples directly two levels of variability and the case bootstrap which is considered to preserve both of them in all evaluated designs. On the other hand, the bootstrap methods which resample only the residuals performed poorly with a large underestimation of the SE of parameters and poor estimates of coverage rates. This is because the between-subject variability of parameters in the evaluated designs were not taken into account. The worse performance was also obtained for the bootstraps combining case and residual, in which the residual variability is considered already resampled in the case bootstrap.

In order to understand the use of bootstrap in the field of population PKPD, we conducted a literature research in PUBMED with the keywords "bootstrap AND population AND (pharmacokinetics OR pharmacodynamics OR pharmacokinetic-pharmacodynamic OR pharmacokinetic/pharmacodynamic) AND (NONMEM OR MONOLIX OR nonlinear mixed effects)" and recovered 90 papers up to November 2012. All of them used the case bootstrap as a model evaluation tool, and only one of them used the parametric bootstrap, as a complement to the case bootstrap. The purpose of using bootstrap in PKPD was mainly for comparing the parameter estimates obtained from the bootstrap datasets with those obtained in the original dataset and estimating SE and/or constructing CI (in 90% papers). It was less frequently used for covariate selection (in 6.7% papers) and structural model selection (in 3.3% papers). Both nonparametric case bootstrap and parametric bootstrap are implemented in programs, Wings for NONMEM and Pearl-speaks-NONMEM (PsN) [13, 14]. While the case bootstrap has been mostly used in population PKPD, there have been very few studies in the literature to evaluate its performance, especially in comparison with the nonparametric/parametric random effect and residual bootstrap which may better approach the "true" data generating process.

In the present paper, we evaluated the performance different bootstrap methods by a simulation study and compared them to that of the asymptotic method in estimating uncertainty of parameters in NLMEM with heteroscedastic error. This simulation design was based on real PK data collected from two clinical trials of aflibercept, an anti-VEGF drug, in cancer patients.

2 METHODS

2.1 Statistical models

Let y_{ij} denote the observation j of subject i at time t_{ij} , where $i = 1, \dots, N$; $j = 1, \dots, n_i$, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ regroups the $(n_i \times 1)$ vector of measurements in subject i , $\xi_i = (t_{i1}, t_{i2}, \dots, t_{in_i})^T$ presents the design vector of subject i , $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ regroups all the measurements from N subjects, $n_{tot} = \sum_{i=1}^N n_i$ denotes the total number of observations. We define an NLMEM as follows:

$$\begin{cases} \mathbf{y}_i = f(\xi_i, \phi_i) + g(\xi_i, \phi_i, \sigma)\varepsilon_i \\ \phi_i = h(\mu, \eta_i) \\ \eta_i \sim N(0, \Omega) \\ \varepsilon_i \sim N(0, 1) \end{cases} \quad (1)$$

where f is the structural model, g is the residual error model, σ is the vector of parameters in the residual error model, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})$ are normally distributed errors with mean zero and variance 1, ϕ_i is the $(p \times 1)$ vector of individual regression parameters, μ is the $(p \times 1)$ vector of fixed effects, h is the function of individual parameters ϕ_i , η_i is the $(q \times 1)$ vector containing the random effects and Ω is the $(q \times q)$ covariance matrix of the random effects. The random effects η_i and the residual errors ε_i are assumed to be independent for different subjects and to be independent of each other for the same subject. The individual PK parameters are often assumed to follow log-normal distribution $\phi_i = \mu \exp(\eta_i)$. Special cases for g include $g = \sigma_a$ (constant/homoscedastic error model) and $g = \sigma_p f(\xi_i, \phi_i)$ (proportional error model).

2.2 Estimation methods

The parameters of NLMEM are estimated by maximizing the log-likelihood function $L(\mathbf{y}, \theta)$ of the response \mathbf{y} , with $\theta = (\mu, \Omega, \sigma)$ the $(l \times 1)$ vector of all parameters of the model. This function is given by:

$$\begin{aligned} L(\mathbf{y}; \theta) &= \sum_{i=1}^N L(\mathbf{y}_i; \theta) = \sum_{i=1}^N \log \left(\int p(\mathbf{y}_i, \phi_i; \theta) d\phi_i \right) \\ &= \sum_{i=1}^N \log \left(\int p(\mathbf{y}_i | \phi_i; \theta) p(\phi_i; \theta) d\phi_i \right) \end{aligned} \quad (2)$$

where $p(\mathbf{y}_i | \phi_i; \theta)$ is the conditional density of the observations given the random effects, $p(\phi_i; \theta)$ is the density of the individual parameters and $p(\mathbf{y}_i, \phi_i; \theta)$ is the likelihood of the complete data (\mathbf{y}_i, ϕ_i) of subject i .

As the random effects are unobservable and the regression function is nonlinear, the likelihood of NLMEM has no closed form and cannot not be evaluated analytically. The likelihood is usually approximated by linearisation of the function f , such as First Order (FO) and First Order Conditional Estimation (FOCE) methods implemented in NONMEM [15]. These methods linearise the structural model either around the expectation of the random effects (FO) or around the individual predictions of the random effects (FOCE). Although the linearisation methods are numerically efficient, they have the potential of producing inconsistent estimates when the between-subject variability is high and/or the number of observations per subject increases slower than the number of subjects [16, 17]. An alternative method to linearisation is to use the Stochastic Approximation Expectation-Maximization (SAEM) algorithm as an exact ML computation [18]. This algorithm consists, at each iteration, in successively simulating the random effects with the conditional distribution (E step) using Markov Chain Monte-Carlo procedure and updating the unknown parameters of the model (M step). In this simulation study, we used the SAEM algorithm implemented in MONOLIX 4.1.2 (Matlab version) as the estimation method.

The ML estimate $\hat{\theta}$ of θ is asymptotically normally distributed with mean θ and asymptotic estimation covariance matrix given by the inverse of \mathbf{M}_F . \mathbf{M}_F is computed as the negative Hessian of the loglikelihood in all the model parameters:

$$\mathbf{M}_F = \sum_{i=1}^N \frac{-\partial^2 L(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

As the likelihood has no closed form, linearisation of the model around the conditional expectation of the individual parameters has been proposed to derive an approximate expression of \mathbf{M}_F . This approach is implemented in MONOLIX and used in this simulation study.

The asymptotic SE of parameters are then estimated as the square root of the diagonal element of the estimated covariance matrix.

When the parameters of the model have been estimated, empirical Bayes estimates (EBEs) of the individual parameters ϕ_i can be obtained as the mode of the posterior distribution of ϕ_i :

$$m(\phi_i | \mathbf{y}_i; \hat{\boldsymbol{\theta}}) = \text{Argmax}_{\phi_i} p(\phi_i | \mathbf{y}_i; \hat{\boldsymbol{\theta}})$$

2.3 Bootstrap methods

The principle of the bootstrap is to repeatedly generate pseudo datasets by resampling with replacement from the original sample. The unknown original distribution of parameters may be replaced by the empirical distribution of the sample, which refers to the nonparametric bootstrap [19] or simulated from a parametric distribution, which refers to the parametric bootstrap. In this study, we are interested in bootstrap methods in regression.

Let B be the number of bootstrap samples to be drawn from the original dataset, a general bootstrap algorithm in regression is:

1. Generate a bootstrap sample by resampling from the data and/or from the estimated model
2. Obtain the estimates for all parameters of the model for the bootstrap sample
3. Repeat steps 1 & 2 B times to obtain the bootstrap distribution of parameter estimates and then compute mean, standard deviation, and 95% CI of this distribution as described below

Let $\hat{\boldsymbol{\theta}}_b^*$ be the vector of parameters estimated for the b^{th} bootstrap sample. The bootstrap parameter estimate $\hat{\boldsymbol{\theta}}_B$ is calculated as the average of the parameter estimates from the B bootstrap samples:

$$\hat{\boldsymbol{\theta}}_B = \frac{1}{B} \sum_{b=1}^B (\hat{\boldsymbol{\theta}}_b^*) \quad (3)$$

The bootstrap standard error of the l^{th} component of $\hat{\boldsymbol{\theta}}_B$ is obtained as the standard deviation of the estimated parameters from the B bootstrap samples:

$$\widehat{SE}_B^{(l)} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\boldsymbol{\theta}}_b^{*(l)} - \hat{\boldsymbol{\theta}}_B^{(l)})^2} \quad (4)$$

A 95% bootstrap confidence interval of the l^{th} component of $\boldsymbol{\theta}$ can be constructed by calculating the 2.5th and 97.5th percentiles of the bootstrap distribution. The bootstrap samples are first sorted into ascending order and these percentiles are respectively given by the $(B+1)\alpha^{th}$ and $(B+1)(1-\alpha)^{th}$ elements of the ordered bootstrap samples where $\alpha=0.025$. When $(B+1)\alpha$ does not equal a whole number, interpolation must be used [20]. An alternative approach is to use a normal approximation to construct a bootstrap confidence interval, using the estimated \widehat{SE}_B :

$$[\hat{\boldsymbol{\theta}}_B^{(l)} - \widehat{SE}_B^{(l)} \cdot z_{1-\alpha/2}; \hat{\boldsymbol{\theta}}_B^{(l)} + \widehat{SE}_B^{(l)} \cdot z_{1-\alpha/2}] \quad (5)$$

$z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution ($z_{0.975} = 1.96$). However, it is preferable to use the bootstrap percentile confidence interval when bootstrapping [5, 20].

In the present study, we evaluated the three bootstrap methods which showed a good performance in LMEM [12]: the case bootstrap $B_{\text{case,none}}$, the nonparametric bootstrap of random effects and global residuals $B_{\eta,\text{GR}}$ and the parametric bootstrap $B_{\text{P}\eta,\text{PR}}$. The individual residual bootstrap was not evaluated in this study because this method appeared not to be consistent with the non-correlated structure of residuals, although it had provided similar results to the global residual bootstraps in the LMEM. All the bootstrap methods which do not perform well in LMEM were not evaluated in this NLMEM study.

The detailed algorithms of the evaluated bootstrap methods to obtain a bootstrap sample (bootstrap generating process) are presented below.

2.3.1 Case bootstrap ($B_{\text{case,none}}$)

This method consists of resampling with replacement the entire subjects, that is the joint vector of design variables and corresponding responses (ξ_i, \mathbf{y}_i) from the original data before modeling. It is also called the *paired bootstrap*. It is the most obvious way to do bootstrapping and makes no assumptions on the model.

2.3.2 Nonparametric random effect and residual bootstrap ($B_{\eta,\text{GR}}$)

This method consists of resampling with replacement the random effects obtained after model fitting, as well as the residuals globally. The bootstrap sample is obtained as follows:

1. Fit the model to the data then estimate the random effects $\hat{\eta}_i$ from $\{\hat{\phi}_i\}$ and the standardized residuals $\hat{\epsilon}_{ij} = (\mathbf{y}_i - f(t_{ij}, \hat{\phi}_i)) / g(t_{ij}, \hat{\phi}_i, \hat{\sigma})$
2. Draw a sample $\{\eta_i^*\}$ of size N with replacement from $\{\hat{\eta}_i\}$ by assigning an equal probability $\frac{1}{N}$ to each value
3. Draw a sample $\{\epsilon^*\} = \{\hat{\epsilon}_{i^*j^*}\}$ of size n_{tot} with replacement globally from $\{\hat{\epsilon}_{ij}\}$
4. Generate the bootstrap responses $\mathbf{y}_i^* = f(\xi_i, \hat{\mu}, \eta_i^*) + g(\xi_i, \hat{\mu}, \eta_i^*, \hat{\sigma})\epsilon_i^*$

Note that in mixed-effects modeling, the raw random effects and residuals do not necessarily have a zero mean, and their variance/ covariance matrix does not match the model-estimated residual variance/covariance matrices. That's why the nonparametric bootstrap of the raw residuals yields downwardly biased variance parameter estimates [5,6,21]. Therefore, they both must be rescaled by being centred and then transformed to have empirical variance/covariance matrices equal to those estimated by the model. In linear mixed-effects models, Carpenter et al proposed to center the random effects and residuals and then transform them using correction matrices accounting for the differences between their corresponding estimated and empirical variance-covariance matrices (shrinkage) [21,22]. The correction matrices were calculated using the Cholesky decomposition of both estimated and empirical variance-covariance matrices.

In this study, we extended the correction of Carpenter et al for NLMEM by using Eigen Value Decomposition (EVD), a special case of Singular Value Decomposition (SVD) for square symmetric matrices [23], instead of the Cholesky decomposition. This extension was done to deal with the numeric problems sometimes seen in nonlinear models where some eigenvalues of the variance-covariance matrix are very close to zero. The detailed transformation of random effects and residuals is presented in Appendix.

2.3.3 Parametric random effect and residual bootstrap ($B_{\text{P}\eta,\text{PR}}$)

This methods resamples both random effects and residuals by simulating from estimated distribution after model fitting. The bootstrap sample is obtained as follows:

1. Fit the model to the data
2. Draw a sample $\{\eta_i^*\}$ of size N from a multivariate normal distribution with mean zero and covariance matrix $\hat{\Omega}$
3. Draw N samples $\{\epsilon_i^*\}$ of size n_i from a normal distribution with mean zero and variance one
4. Generate the bootstrap responses $\mathbf{y}_i^* = f(\xi_i, \hat{\mu}, \eta_i^*) + g(\xi_i, \hat{\mu}, \eta_i^*, \hat{\sigma})\epsilon_i^*$

Of note that, the simulation of random effects from a multivariate normal distribution can raise problem when $\hat{\Omega}$ contains one or some eigenvalues close to zero. We proposed to construct pseudo matrix of $\hat{\Omega}$ using EVD: $\hat{\Omega}' = V_{\hat{\Omega}} D_{\hat{\Omega}}' V_{\hat{\Omega}}^T$ where $V_{\hat{\Omega}}$ is the orthogonal matrix resulting from EVD of $\hat{\Omega}$, $D_{\hat{\Omega}}'$ is the diagonal matrix containing eigenvalues of $\hat{\Omega}$ in diagonal entries, of which eigenvalues smaller than a tolerance (10^{-6}) was set to zero.

2.4 Bootstrap methods with stratification

The nonparametric bootstrap methods described above preserve the structure and the characteristics of the original data when the data is homogenous. In the case of unbalanced designs, different groups in the original data should be defined and resampling in each group should be done to maintain a similar structure of the original data in the bootstrap sample [14]. For example when a study includes different numbers of observations in each subject, the case bootstrap will generate the bootstrap samples having different total number of observations. The nonparametric residual bootstrap preserves the same structure of the original data but does not take in to account the different shrinkages of random effects and residuals for groups with different designs. Bootstrap with stratification can be done for these two methods in this example as follows:

Case stratified bootstrap ($B_{case,none}^{strat}$). This method consists in resampling the entire subjects in each group.

Nonparametric residual stratified bootstrap ($B_{\eta,GR}^{strat}$). This method applies the correction for shrinkage and bootstrapping the random effects and the residuals separately in each group. An example of the correction for shrinkage with two data groups with different designs is presented in Appendix.

3 SIMULATION STUDIES

3.1 Motivating example

As an illustrative example, data from two clinical trials of aflibercept, an anti-angiogenic drug for cancer patients were used. The first trial was a phase I dose-escalation study of aflibercept in combination with docetaxel and cisplatin in patients with advanced solid tumors [24]. The second trial was a randomised controlled phase III study of aflibercept and docetaxel versus docetaxel alone after platinum failure in patients with advanced or metastatic non-small-cell lung cancer [25]. Aflibercept was administered intravenously every 3 weeks in combination with docetaxel at dose levels ranging from 2 to 9 mg/kg in the phase I trial and at dose of 6 mg/kg in the phase III trial. In the phase I trial, blood samples were collected at 1, 2, 4, 8, 24, 48, 168, 336 hours after the start of aflibercept administration and before the administration of all subsequent cycles. In the second trial, blood samples were taken pre-dose and at the end of aflibercept infusion on day 1 (cycle 1) and every odd cycles before treatment administration and at approximately 30 and 90 days after the last aflibercept treatment. The free plasma aflibercept concentrations were measured in all samples using enzyme-linked immunosorbent assay (Elisa) method. The limit of quantification (LOQ) for free aflibercept in plasma was 15.6 ng/ml and data below LOQ for both studies (6.3% for the phase I trial and 9.1% for the phase III trial) were omitted.

For this simulation study, we focused mainly on the data after the first dose of the phase I trial and the data after the two first doses of the phase III trial disregarding potential interoccasion variability. All the patients having at least two observations were included in this analysis. This subset contains 344 patients including 53 patients from the phase I trial with an average of 9 observations per patient (rich design group) and 291 patients from the phase III trial with 2 observations per patient (sparse design group).

To describe the PK of free aflibercept, we analysed jointly the data of all patients using MONOLIX 4.1.2 (Matlab version) and the SAEM algorithm for estimating parameters. A two-compartment infusion PK model with first-order linear elimination described the aflibercept concentrations in this subset. The between-subject variability was modeled using an exponential model. The residual variability was chosen among the additive, proportional or combined models using the log-likelihood (LL) test. We examined the SAEM convergence graph and the goodness-of-fit plots to evaluate the chosen model.

The proportional error model was the best residual model. With respect to the model of random effects, there was a high correlation between CL and Q (0.9), which decreased 47 point in -2LL, compared to the model without this correlation. The variability of V_2 was small so it was fixed to zero to have a better convergence of parameters. This did not change significantly the log-likelihood. The parameter estimates of this model are presented in first column of Table 3. All the parameters were estimated with good precision ($RSE < 15\%$) obtained by the asymptotic \mathbf{M}_F method. With the chosen model, the goodness-of-fit was satisfactory.

3.2 Simulation settings

In this simulation, we aimed to evaluate the performance of bootstrap in different designs using the first-order elimination PK model and their parameter estimates developed in the real data (section 3.1): a frequent sampling design, a sparse sampling design and an unbalanced design with mixing of the frequent and the sparse observation designs during resampling. We also aimed to evaluate the bootstraps in models with higher nonlinearity; a mixed-order (Michaelis-Menten) elimination model was used to illustrate this case.

For the balanced designs, the sampling times were identical for all subjects.

First-order elimination frequent sampling design. We simulated $N=30$ subjects with $n=9$ observations per subject at 1, 2, 4, 8, 24, 48, 168, 336, 503 hours after administration of aflibercept. In this design, we used the same model as the real data, except for a smaller correlation between CL and Q (0.5 instead of 0.9) to avoid the convergence problems.

First-order elimination sparse sampling design. We simulated $N=70$ subjects with $n=4$ observations per subject at 1, 24, 48, 503 hours after administration of aflibercept. In this design, we removed the correlation between CL and Q as well as the variability for Q to avoid the convergence problems. The variability of other parameters was set to 30%.

Mixed-order (Michaelis-Menten) elimination frequent sampling design. We simulated $N=30$ subjects with $n=9$ observations per subject at 1, 2, 4, 8, 24, 48, 168, 336, 503 hours after administration of aflibercept. In this design, we used the same model of the linear sparse design but replaced first-order elimination ($CL = 0.04(l/h)$) by mixed-order elimination ($V_{max} = 2(mg/h)$ with $\omega_{V_{max}} = 30\%$ and $K_m = 20mg/l$ with $\omega_{K_m} = 0\%$) to increase the nonlinearity of the model. The values of V_{max} and K_m were chosen based on the derivatives of the models with respect to these parameters.

For the unbalanced design, we used the first-order elimination model in the sparse balanced design and simulated two groups of patients: $N_1=15$ with $n_1=9$ observations per subject at 1, 2, 4, 8, 24, 48, 168, 336, 503 hours after administration of aflibercept, and $N_2=75$ with $n_2=2$ observations per subject at 1, 503 hours after administration of aflibercept. The ratio of patients with frequent and sparse sampling was similar to that in the real data (16.7% subjects have rich sampling times with an average of 9 observations and 83.3% subjects have only 2 observations).

All the designs had approximately the same total number of observations (~ 280 observations).

For each design, we simulated $K=100$ replications. The SAEM algorithm implemented in the MONOLIX 4.1.2 was used to fit the data. The asymptotic SE of parameters were obtained by the inverse of \mathbf{M}_F , computed as the negative Hessian of loglikelihood (described previously in Methods section). The bootstrap algorithms were implemented in R 2.14.1. All the bootstrap datasets were fitted with the initial values obtained from estimates from the original data.

Examples of simulated data for each given design are illustrated in Figure 1.

3.3 Evaluation of bootstrap methods

We drew $B=999$ bootstrap samples for each replication of simulated data and for each bootstrap method. $B=999$ was chosen to directly estimate the quantiles for 95% CI without interpolation [5,20]. For each method, we therefore performed 99900 fits (100 simulated datasets \times 999 bootstrap datasets).

For the k^{th} simulated dataset and for a given bootstrap method, we computed the bootstrap parameter estimate $\hat{\theta}_{B;k}^{(l)}$ as in equation 3, using 999 bootstrap samples, and the bootstrap SE estimate $\widehat{SE}_{B;k}^{(l)}$ as in equation 4 as well the CI, for the l^{th} component of θ . The relative bias (RBias) of the l^{th} component of bootstrap estimate $\hat{\theta}_B$ was obtained by comparing the bootstrap estimate $\hat{\theta}_{B;k}^{(l)}$ and the estimate $\hat{\theta}_k^{(l)}$ as follows:

$$RBias(\hat{\theta}_B^{(l)}) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{\theta}_{B;k}^{(l)} - \hat{\theta}_k^{(l)}}{\hat{\theta}_k^{(l)}} \times 100 \right) \quad (6)$$

The average bootstrap SE was obtained by averaging the SE from equation (6) over the $K=100$ datasets. The true SE is unknown, but we can get an empirical estimate as the standard deviation of the differences between the estimate of the parameter in the K datasets and the true value θ_0 :

$$\widehat{SE}_{\text{empirical}}^{(l)} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_k^{(l)} - \theta_0^{(l)})^2} \quad (7)$$

The relative bias on SE of the l^{th} component of $\hat{\theta}_B$ was then obtained by comparing the average SE to the empirical SE:

$$RBias(\widehat{SE}_B^{(l)}) = \frac{\frac{1}{K} \sum_{k=1}^K \widehat{SE}_{B;k}^{(l)} - \widehat{SE}_{\text{empirical}}^{(l)}}{\widehat{SE}_{\text{empirical}}^{(l)}} \times 100 \quad (8)$$

The coverage rate of the 95% bootstrap confidence interval (CI) was defined as the percentage of the $K=100$ datasets in which the bootstrap CI contains the true value of the parameter.

The bootstrap approaches were compared in terms of the Rbias on the bootstrap parameter estimates, the Rbias on SE, and the coverage rate of the 95% CI of all parameter estimates from all bootstrap samples. The performance of the bootstrap methods were also compared to the performance of the asymptotic method. The relative bias of the l^{th} component of the estimate $\hat{\theta}$ was obtained by comparing the estimate $\hat{\theta}_k^{(l)}$ and the true value $\theta_0^{(l)}$ as follows:

$$RBias(\hat{\theta}^{(l)}) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{\theta}_k^{(l)} - \theta_0^{(l)}}{\theta_0^{(l)}} \times 100 \right) \quad (9)$$

The relative bias of asymptotic SE estimate was defined in the same way as equation (8), but with respect to $\widehat{SE}_k^{(l)}$ being the asymptotic SE of $\hat{\theta}_k^{(l)}$ instead of $\widehat{SE}_{B;k}^{(l)}$. The coverage rate of the 95% asymptotic CI was defined as the percentage of datasets in which the asymptotic CI contains the true value of the parameter.

The bootstrap parameters estimates and their SE were defined as unbiased when relative bias was within $\pm 10\%$, moderately biased (relative bias from $\pm 10\%$ to $\pm 20\%$) or strongly biased (relative bias $> \pm 20\%$). The coverage rate of the 95% CI was considered to be good (from 90% to 100%), low (from 80% to 90%) or poor ($< 80\%$). A good bootstrap was defined as a method providing unbiased estimates for the parameters and their corresponding SE, and ensuring a good coverage rate of the 95% CI. The same criteria were used to evaluate the performance of the asymptotic method.

3.4 Application to real data

All the bootstrap methods evaluated in the simulations studies were then applied to the real data with $B=999$ replications. The parameter and SE estimates obtained by the bootstrap methods were compared to those obtained by the asymptotic approach.

4 RESULTS

4.1 Simulation studies

Performance of bootstrap in balanced designs

The performance of the three bootstrap methods as well as the asymptotic method regarding the relative bias of parameters, their SE and the coverage rate of 95% CI are presented in Figure 2 and Table 1. In the frequent sampling setting with first-order elimination, all bootstrap methods showed essentially no bias for all parameters. In terms of SE estimation, the case bootstrap ($B_{\text{case},\text{none}}$) yielded moderate bias for SE of Q (22.1%); the nonparametric random effect and residual bootstrap ($B_{\eta,\text{GR}}$) showed a small bias for SE of ω_Q and σ_p ($< 13.2\%$) while the parametric random effect and residual bootstrap ($B_{p\eta,\text{PR}}$) estimated correctly all the SE. In terms of coverage rate, all the bootstrap methods provide good coverage rate for all parameters, except for low coverage rate of the correlation between CL and Q (ρ) (83%) observed for $B_{\eta,\text{GR}}$. The asymptotic method performed however less well than the bootstrap methods with a greater bias for ρ and SE of several parameters (V_1 , Q , ω_Q , ρ) and poorer coverage rates for V_1 , Q , and ρ .

In the first-order elimination sparse sampling setting where the correlation between CL and Q was omitted and the variability of parameters was set to 30%, the $B_{\text{case},\text{none}}$ and the $B_{p\eta,\text{PR}}$ showed no bias in parameter estimates, SE and provided good coverage rates for all parameters except for a large bias for SE of Q ($> 100\%$). The $B_{\eta,\text{GR}}$ not only provided large bias for SE of Q , but also gave a bias for SE of ω_{V_1} and poor coverage rates for ω_{V_1} and ω_{V_2} . The asymptotic method, however, performed very well and better than the bootstrap methods.

Higher nonlinearity was evaluated in the mixed-order elimination frequent sampling setting. In this design, all bootstrap methods estimate correctly all the parameters except for a moderate bias on ω_{V_2} observed for $B_{\eta,\text{GR}}$. In terms of SE estimation, the bootstrap methods provided good estimates for the SE of all parameters, with the exception for underestimation of SE of K_m observed for all methods (-17.6 to -13.6%). In addition, the $B_{\text{case},\text{none}}$ underestimated SE of σ_p and $B_{\eta,\text{GR}}$ underestimated SE of ω_{V_2} . In terms of coverage rate, the bootstrap methods provided good coverage rates for first-order elimination parameters but gave poor to low coverage rates for more highly nonlinear parameters (87% for V_{max} and 71-77% for K_m). Compared to the bootstrap methods, the asymptotic method showed higher bias for SE of almost all parameters, especially for nonlinear parameters, e.g K_m (-89.3%); which leads to extremely poor coverage rate for K_m (19%). The asymptotic method also provided poor coverage rate for Q (77%).

The boxplots of the relative error of SE of all parameter estimates obtained by the bootstrap methods and the asymptotic method in all evaluated designs are shown in Figure 3. The range of relative errors of bootstrap SE across the $K=100$ replications did not show any practically relevant differences across the bootstrap methods. The range of relative errors was however different between the asymptotic method and the bootstrap methods. In the sparse sampling design, the estimates of SE obtained by the asymptotic method were more accurate and precise, especially for Q . On the contrary, the asymptotic method performed less well than the bootstrap methods in frequent sampling designs with underestimation of SE of several parameters, especially SE of K_m in the mixed-order design where SE of V_2 and V_{max} were also overestimated.

Performance of bootstrap in the unbalanced design

In the fourth simulation setting, we considered a highly unbalanced design where 16.7% subjects have rich sampling times with average 9 observations and 83.3% subjects have only 2 observations. Figure 4 and Table 2 present the results of the bootstrap methods with and without stratification and the asymptotic method in this simulation. All the non-stratified bootstrap methods estimated correctly the parameters, except for a slight overestimation of ω_{V_2} observed in the $B_{p\eta,\text{PR}}$. In terms of SE estimation, the case bootstrap did not estimate well the SE of Q with a bias of 34.3%; there were also a high bias on SE of ω_{V_1} (40.8%), ω_{CL} (20%), σ_p (21.1%) for the $B_{\eta,\text{GR}}$. In terms of coverage rates, all the bootstrap methods provided good coverage rates, except for a lower coverage rate of σ_p in $B_{\text{case},\text{none}}$ and $B_{p\eta,\text{PR}}$. The asymptotic method performed reasonably well in this unbalanced design with only a small bias on SE of V_2 and lower coverage rate of Q .

The stratified bootstrap methods were also evaluated in this unbalanced design in order to maintain the same structure of the original dataset. The stratified case bootstrap $B_{\text{case},\text{none}}^{\text{strat}}$ reduced the bias in SE of Q

but this bias was still high (21.7 vs 34.3%). The stratified residual bootstrap $B_{\eta,GR}^{strat}$ reduced the bias in SE of ω_{V_1} and σ_p but overestimated SE of Q and gave lower coverage rates for almost parameters compared to the non-stratified version.

Figure 5 presents the boxplots of the relative error of SE of all parameter estimates obtained by the bootstrap methods and the asymptotic method in the unbalanced design. The $B_{case,none}$ and $B_{case,none}^{strat}$ estimated with less precision the SE of Q with a very large variability compared to other bootstrap methods. The $B_{P\eta,PR}$ and the asymptotic method estimate most correctly and precisely the SE of all parameter estimates.

Of note that, 100% simulated and bootstrap datasets converged in all evaluated designs.

4.2 Application to real data

Table 3 presents the median of the parameter estimates obtained by the bootstrap methods for the real dataset and the bootstrap relative standard errors with respect to the original parameter estimates. We found that all the bootstrap methods had similar medians of parameter estimates, except for Q with the difference between $B_{P\eta,PR}$ and the other methods. In terms of precision estimation, the bootstrap methods provided good RSE for all parameters (<24%). However, there were some differences for the estimation of RSE of Q and variance parameters. Similar to the simulation results in the unbalanced design, $B_{\eta,GR}$ and $B_{\eta,GR}^{strat}$ gave higher RSE for the variance parameters compared to the other methods. The asymptotic method and $B_{P\eta,PR}$ had very similar estimates for almost all parameters in terms of both parameter and RSE estimation, including RSE of σ_p which was larger for other methods.

The bootstrap confidence intervals for each parameter are shown in Figure 6. The parameter estimates of the real dataset were contained within the CI obtained by all bootstrap methods, with the exception of σ_p which lay outside the CI of $B_{\eta,GR}$ and $B_{\eta,GR}^{strat}$. For $B_{\eta,GR}^{strat}$, the estimates of V_1 , Q , ω_{V_1} were located on the boundary of the bootstrap CI. Compared to the asymptotic CI, the CI of $B_{case,none}$ and $B_{case,none}^{strat}$ for all parameters were similar except for Q while the CI of $B_{\eta,GR}$ and $B_{\eta,GR}^{strat}$ were different, especially for the stratified version.

5 DISCUSSION

In the present paper, we evaluated the performance of the case bootstrap and the nonparametric/parametric bootstrap of random effects and residuals by a simulation study and compared them to that of the asymptotic method in estimating uncertainty of parameters in NLMEM with heteroscedastic error. This simulation was based on a real PK data collected from two clinical trials of aflibercept, a novel anti-angiogenic drug, in cancer patients.

When dealing with NLMEM, we should consider other factors which influence the bootstrap, such as the nonlinearity of the model and the heteroscedasticity. It should be noted that bootstrapping nonlinear models is done in the same manner as bootstrapping linear models. However, the nonlinearity makes the estimation process much more difficult and laborious. The bootstrap becomes time consuming because we need to perform a nonlinear estimation for each bootstrap sample. The complexity increases when the residual error model is heteroscedastic (the variance of residuals errors is not constant) because the residuals can not be interchangeable. The algorithm for bootstrapping the residuals will not be valid because the bootstrapped dataset might not have the same variance model as the original data. To overcome this issue, the residuals errors need to be standardized to have the same variance before bootstrapping [26]. However, heteroscedasticity is not a problem for the case bootstrap because heteroscedasticity will be preserved after bootstrapping.

Another issue which is very important when applying the nonparametric residual bootstrap in NLMEM is the transformation of the raw residuals to avoid the underestimation in variance parameter estimates [5, 6, 21]. The shrinkage correction using the ratio matrix between the empirical and the estimated variance covariance matrices was proposed by Carpenter et al, using the Cholesky decomposition for a positive definite matrix [21, 22]. This correction performed very well for LMEM [12]. In NLMEM, the numerical problems make the empirical and/or estimated variance-covariance matrices sometimes closer to a semi-positive definite matrix with one or several eigenvalues close to zero. We used the EVD, a special case of

SVD for square symmetric matrices [23], to obtain the ratio matrix by creating the pseudoinverse matrices if the variance-covariance matrices are not strictly positive definite.

Our simulation study evaluated the bootstrap methods in both balanced and unbalanced designs, with first-order or mixed-order elimination PK models, representing low and higher degrees of nonlinearity. In the frequent sampling balanced designs with first-order or mixed-order elimination with a small number of patients (30 subjects), the studied bootstraps improved the description of uncertainty of some parameters compared to the asymptotic method, particularly for parameters which enter the model most non-linearly such as V_{max} and K_m . The case bootstrap and the parametric bootstrap performed similarly, except for a higher bias on SE of Q observed for the case bootstrap. The nonparametric bootstrap of random effects and residuals, however, performed less well with higher bias for some variance parameters and their SE, leading to poorer coverage rates for these parameters. In the sparse sampling design with first-order elimination, the bootstrap methods performed less well than the asymptotic method because they yielded very high bias for SE of Q (>100%). This may due to the skewed distributions of estimates of Q obtained by all the bootstrap methods and the sensitivity of bootstrap methods to extreme values. One of strategies for dealing with this problem is to bootstrap with Winsorization by giving less weight to values in the tails of distribution and paying more attention to those near the center [27]. The Winsorization approach set all outliers to a specified percentile of the data before computing the statistics. Note that, it is not equivalent to simply throwing some of the data away. This approach, however, was not evaluated in this study.

Compared to the results in LMEM, the performance of the evaluated bootstrap methods are shown to be more different in NLMEM: the case bootstrap is more sensitive to the skewed distribution in parameter estimates, the nonparametric residual bootstrap yields more bias for the uncertainty of variance parameters while the parametric bootstrap provides the best description of uncertainty of all parameters in a setting where simulation and resampling distributions were identical, but may seems less robust in case of model misspecification.

In the unbalanced design with first-order elimination (containing 83.3% subjects with only 2 observations), the case bootstrap was more sensitive to extreme values, giving highest bias for SE of Q and had poorer coverage rates for σ_p . The stratification on the design with the case bootstrap reduced the bias on SE but it was still high. The nonparametric residual bootstrap was less sensitive to the extreme values, but gave higher bias for SE of variance parameters. As the shrinkages for random effects in the frequent and the sparse sampling groups are different, the global correction of random effects may not be a good solution. We tested the most simple stratification, first in the correction step in which we correct the empirical variance matrix in each group with respect to the estimated variance matrix, and second in the resampling step. This stratification did not improve much the bias on SE of parameters; in addition, it provided low to poor coverage rates of almost all parameters. A better performance was observed for the parametric residual bootstrap and the asymptotic method with only a slight lower coverage rate of σ_p in $B_{P\eta,PR}$ and lower coverage rate of Q for the asymptotic method.

The bootstrap methods were applied to real PK data of aflibercept from two clinical trials in cancer patients. The medians of parameter estimates obtained by all bootstrap methods were generally in agreement with the original parameter estimates using the chosen PK model, except for lower values of Q estimated with almost bootstrap methods. This difference may be related to the high correlation between CL and Q in the original data, which was reduced or omitted in the simulation study. The results of the nonparametric bootstrap of random effects and residuals in the real dataset were similar to the simulation findings in the unbalanced design, the larger SE of variance parameters and the failure of the simple stratification for this method based on rich/sparse design in both correction and resampling steps. Similarly, the case bootstrap with and without stratification provided different confidence intervals for Q , the parameter having the largest RSE, compared to the asymptotic and the parametric bootstrap method. Also, they gave higher RSE for the residual variance, which may be the results of large amount of the sparse data in the real dataset.

In conclusion, our simulation study showed that the asymptotic method performed well in most cases while the bootstrap methods provided better estimates of uncertainty for parameters with high nonlinearity. Overall, the case bootstrap proved the most robust method, and it was also fast and simple as it makes no assumptions on the model and preserves both between subject and residual variability in one resampling step. The nonparametric residual bootstrap also performed well, with a correction for variance underestimation before resampling. However, the bootstrap methods may face several practical problems. They can generate a wrong estimate of SE of a parameter with a skewed distribution when the estimation is poor.

In addition, bootstrapping in unbalanced designs is much more challenging, and stratification may be insufficient to correct for heterogeneity especially in very unbalanced designs. This study gave us a clearer picture about the statistical properties of bootstrap methods in NLMEM for estimating the uncertainty of parameters. However, some issues raised through our results will need to be addressed in further studies, such as bootstrapping in presence of extreme values for providing more robust SE, correction for shrinkage in the unbalanced designs and performance of bootstraps in case of misspecification in model structure or parameter distributions.

Acknowledgements The authors would like to thank Drug Disposition Department, Sanofi, Paris which supported Hoai-Thu Thai by a research grant during this work. We also thank IFR02 and Hervé Le Nagard for the use of the "centre de biomodélisation".

References

1. Sheiner LB, Rosenberg B, Marathe VV (1977) Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm* 5:445–479.
2. Efron B (1979) Bootstrap methods: Another look at the jackknife. *Annal Stat* 7:1–26.
3. Efron B, Tibshirani RJ (1994) *An Introduction to the Bootstrap* Chapman & Hall, New York.
4. Shao J, Tu D (1995) *The Jackknife and Bootstrap* Springer, New York.
5. Davison AC, Hinkley DV (1997) *Bootstrap Methods and their Application* Cambridge University Press, Cambridge.
6. MacKinnon JB (2006) Bootstrap methods in econometrics. *Econ Rec* 82:S2–S18.
7. Wehrens R, Putter H, Buydens LMC (2000) The bootstrap: a tutorial. *Chemom Intell Lab Syst* 54:35–52.
8. Das S, Krishen A (1999) Some bootstrap methods in nonlinear mixed-effect models. *J Stat Plan Inference* 75:237–245.
9. Halimi R (2005) *Nonlinear Mixed-effects Models and Bootstrap resampling: Analysis of Non-normal Repeated Measures in Biostatistical Practice* VDM Verlag, Berlin.
10. Van der Leeden R, Busing FMTA, Meijer E (1997) Bootstrap methods for two-level models. Technical Report PRM 97-04, Leiden University, Department of Psychology, Leiden.
11. Wu H, Zhang JT (2002) The study of long-term hiv dynamics using semi-parametric non-linear mixed-effects models. *Stat Med* 21:3655–3675.
12. Thai HT, Mentre F, Holford NH, Veyrat-Follet C, Comets E (2013) A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharm Stat*. Doi: 10.1002/pst.1561.
13. Parke J, Holford NH, Charles BG (1999) A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models. *Comput Methods Programs Biomed* 59:19–29.
14. Lindbom L, Pihlgren P, Jonsson EN (2005) PsN-Toolkit—A collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed* 79:241–257.
15. Beal S, Sheiner LB (1989) *NONMEM User's Guide-Part i. User Basic Guide.*, (University of California, San Francisco), Technical report.
16. Vonesh EF, Chinchilli VM (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements* Marcel Dekker, New York.
17. Ge Z, Bickel P, Rice J (2004) An approximate likelihood approach to nonlinear mixed effects models via spline approximation. *Comput Stat Data Anal* 46:747–776.
18. Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the em algo. *Annal Stat* 27:94–128.
19. Ette EI (1997) Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol* 37:486–495.
20. Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Stat Med* 19:1141–1164.
21. Carpenter JR, Goldstein H, Rasbash J (2003) A novel bootstrap procedure for assessing the relationship between class size and achievement. *Appl Statist* 52:431–443.
22. Wang J, Carpenter JR, Kepler MA (2006) Using SAS to conduct nonparametric residual bootstrap multilevel modeling with a small number of groups. *Comput Methods Programs Biomed* 82:130–143.
23. Kalman D (1996) A singularly valuable decomposition: The SVD of a matrix. *College Math J* 27:1–23.
24. Freyer G, Isambert N, You B, Zanetta S, Falandry C, Favier L, Trillet-Lenoir V, Assadourian S, Soussan-Lazard K, Ziti-Ljajic S, Fumoleau P (2012) Phase I dose-escalation study of aflibercept in combination with docetaxel and cisplatin in patients with advanced solid tumours. *Br J Cancer* 107:598–603.
25. Ramlau R, Gorbunova V, Ciuleanu TE, Novello S, Ozguroglu M, Goksel T, Baldotto C, Bennouna J, Shepherd FA, Le-Guenneec S, Rey A, Miller V, Thatcher N, Scagliotti G (2012) Aflibercept and docetaxel versus docetaxel alone after platinum failure in patients with advanced or metastatic non-small-cell lung cancer: a randomized, controlled phase III trial. *J Clin Oncol* 30:3640–3647.
26. Bonate PL (2011) *Pharmacokinetic-pharmacodynamic modeling and simulation* Springer, New-York, Second edn.
27. Ette EI, Onyiah LC (2002) Estimating inestimable standard errors in population pharmacokinetic studies: the bootstrap with winsorization. *Eur J Drug Metab Pharmacokinet* 27:213–24.

APPENDIX

Appendix A: Transformations of random effects using Eigen Value Decomposition (EVD)

Let S and $\hat{\Omega}$ denote respectively the square symmetric empirical and estimated variance covariance matrices. Let \hat{U} denote the matrix of rescaled raw estimated random effects (by centering to ensure to have zero mean) with the column entries corresponding to the random effects of each parameter. The empirical matrix S of \hat{U} is then defined as:

$$S = \frac{\hat{U}^T \hat{U}}{N-1} \quad (\text{A.1})$$

where N is the number of subjects

The transformed matrix of random effects \hat{U}' is defined using the correction matrix A_η :

$$\hat{U}' = \hat{U} A_\eta \quad (\text{A.2})$$

The matrix A_η is formed using the Eigen Value Decomposition (EVD) of the matrix S and the matrix $\hat{\Omega}$. Since any square symmetric matrix M can be decomposed by EVD in an expression involving a diagonal matrix (D) containing the eigenvalues of M in diagonal entries and a orthogonal matrix (V) ($VV^T = I$, I is identity matrix) as follows: $M = VDV^T$, we can write the EVD decomposition of S and $\hat{\Omega}$ as:

$$S = V_S D_S V_S^T \quad (\text{A.3})$$

$$\hat{\Omega} = V_{\hat{\Omega}} D_{\hat{\Omega}} V_{\hat{\Omega}}^T \quad (\text{A.4})$$

where $V_{\hat{\Omega}}$ and V_S are respectively the orthogonal matrices of $\hat{\Omega}$ and S ; $D_{\hat{\Omega}}$ and D_S are respectively the diagonal matrices of $\hat{\Omega}$ and S , where their diagonal entries are respectively the eigenvalues of $\hat{\Omega}$ and S .

We propose the following solution of A which makes the variance covariance matrix of the transformed matrix of random effects \hat{U}' equal to $\hat{\Omega}$:

$$A_\eta = V_S D_S^{-1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T \quad (\text{A.5})$$

where $D_{\hat{\Omega}}^{1/2}$ and $D_S^{-1/2}$ are respectively the diagonal matrices where their diagonal entries are respectively the root square of eigenvalues of $\hat{\Omega}$ and the inverse of root square of eigenvalues of S .

The transposed form of A is:

$$A_\eta^T = V_{\hat{\Omega}} D_{\hat{\Omega}}^{1/2} D_S^{-1/2} V_S^T \quad (\text{A.6})$$

Indeed, the variance covariance matrix of \hat{U}' is equal to $\hat{\Omega}$, based on the following development:

$$\begin{aligned} \frac{\hat{U}'^T \hat{U}'}{N-1} &= \frac{(\hat{U} A_\eta)^T \hat{U} A_\eta}{N-1} = \frac{A_\eta^T \hat{U}^T \hat{U} A_\eta}{N-1} = A_\eta^T S A_\eta \\ &= V_{\hat{\Omega}} D_{\hat{\Omega}}^{1/2} D_S^{-1/2} V_S^T V_S D_S V_S^T V_S D_S^{-1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T \\ &= V_{\hat{\Omega}} D_{\hat{\Omega}}^{1/2} [D_S^{-1/2} (V_S^T V_S) D_S (V_S^T V_S) D_S^{-1/2}] D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T \\ &= V_{\hat{\Omega}} D_{\hat{\Omega}}^{1/2} [D_S^{-1/2} D_S D_S^{-1/2}] D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T \\ &= V_{\hat{\Omega}} D_{\hat{\Omega}}^{1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T \\ &= V_{\hat{\Omega}} D_{\hat{\Omega}} V_{\hat{\Omega}}^T \\ &= \hat{\Omega} \end{aligned} \quad (\text{A.7})$$

In case $D_S^{-1/2}$ is not invertible (singular) due to some eigenvalues of D_S are close to zero, a pseudo-inverse matrix is used by inverting only the positive eigenvalues greater than a tolerance of 10^{-6} and setting those lower to zero.

In balanced designs, the transformation of random effects was carried out in the following steps:

1. Center the raw estimated random effects: $\tilde{\eta}_i = \hat{\eta}_i - \bar{\eta}$
2. Calculate the correction matrix A_η . Let $\hat{\Omega}$ be the model estimated variance-covariance matrix of random effects and S denote the empirical variance-covariance matrix of the centered random effects. The correction matrix is formed using EVD of these two matrices: $A_\eta = V_S D_S^{-1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T$
3. Transform the centered random effects using the ratio A_η : $\hat{\eta}'_i = \tilde{\eta}_i A_\eta$

In unbalanced designs, for example with two groups containing rich and sparse data, the transformation of random effects was carried out in the following steps:

1. Center the raw estimated random effects: $\tilde{\eta}_i = \hat{\eta}_i - \bar{\eta}$ and divide them into two groups $\tilde{\eta}_{G_1}$ and $\tilde{\eta}_{G_2}$ presenting the centered random effects in group 1 (G_1) and group 2 (G_2).
2. Calculate the correction matrices $A_{\eta;G_1}$ and $A_{\eta;G_2}$. Let $\hat{\Omega}$ be the model estimated variance-covariance matrix of random effects, S_1 and S_2 denote the empirical variance-covariance matrix of the centered random effects $\tilde{\eta}_{G_1}$ and $\tilde{\eta}_{G_2}$ respectively. The correction matrices $A_{\eta;G_1}$ and $A_{\eta;G_2}$ using EVD are formed as follows:

$$A_{\eta;G_1} = V_{S_1} D_{S_1}^{-1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T$$

$$A_{\eta;G_2} = V_{S_2} D_{S_2}^{-1/2} D_{\hat{\Omega}}^{1/2} V_{\hat{\Omega}}^T$$

where V_{S_1} and V_{S_2} are respectively the orthogonal matrices of S_1 and S_2 ; D_{S_1} and D_{S_2} are respectively the diagonal matrices of S_1 and S_2 .

3. Transform the centered random effects in each group:

$$\hat{\eta}'_{i;G_1} = \tilde{\eta}_{i;G_1} A_{\eta;G_1}$$

$$\hat{\eta}'_{i;G_2} = \tilde{\eta}_{i;G_2} A_{\eta;G_2}$$

Appendix B: Transformations of residuals

In balanced designs, the transformation of residuals was carried out in the following steps:

1. Center the raw estimated standardized residuals: $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} - \bar{\epsilon}$
2. Calculate the correction factor A_σ

$$A_\sigma = 1/\sigma_{\text{emp}}$$

where σ_{emp} is simply the empirical standard deviation of the raw standardized residuals

3. Transform the centered residuals using the ratio A_σ : $\hat{\epsilon}'_{ij} = \tilde{\epsilon}_{ij} A_\sigma$

In unbalanced designs, for example with two groups containing rich and sparse data, the transformation of residuals was carried out in the following steps:

1. Center the raw estimated standardized residuals: $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} - \bar{\epsilon}$ and divide them into two groups $\tilde{\epsilon}_{G_1}$ and $\tilde{\epsilon}_{G_2}$ presenting the centered residuals in group 1 (G_1) and group 2 (G_2).
2. Calculate the correction factors $A_{\sigma;G_1}$ and $A_{\sigma;G_2}$.

$$A_{\sigma;G_1} = \frac{1}{\sigma_{\text{emp};G_1}}$$

$$A_{\sigma;G_2} = \frac{1}{\sigma_{\text{emp};G_2}}$$

where $\sigma_{\text{emp};G_1}$ and $\sigma_{\text{emp};G_2}$ are respectively the empirical standard deviation of the raw standardized residuals in G_1 and G_2 .

3. Transform the centered residuals in each group:

$$\begin{aligned}\hat{\epsilon}'_{ij;G_1} &= \tilde{\epsilon}_{ij;G_1} A_{\sigma;G_1} \\ \hat{\epsilon}'_{ij;G_2} &= \tilde{\epsilon}_{ij;G_2} A_{\sigma;G_2}\end{aligned}$$

6 TABLES

Table 1 – Relative bias of parameter and standard error (SE) estimates and coverage rate obtained by the asymptotic method (Asym) and the three bootstrap methods for the studied balanced designs

Design	Parameters	Relative bias of parameters (%)				Relative bias of SE (%)				Coverage rate (%)			
		Asym	B _{case,none}	B _{η,GR}	B _{Pη,PR}	Asym	B _{case,none}	B _{η,GR}	B _{Pη,PR}	Asym	B _{case,none}	B _{η,GR}	B _{Pη,PR}
First-order frequent	V_1	-0.0	-0.3	0.3	0.0	-17.5	-8.4	-10.7	-8.9	89	95	93	94
	V_2	-0.6	0.8	-1.4	-0.6	-2.3	8.8	4.0	7.3	94	97	95	97
	Q	-5.9	1.4	-2.6	-2.0	-22.1	22.7	0.7	7.3	87	95	94	94
	CL	-0.6	-0.0	-0.1	-0.1	2.8	4.0	1.5	3.5	95	97	94	97
	ω_{V_1}	-2.4	-4.4	-8.1	-4.8	0.4	8.4	3.3	6.0	97	91	91	93
	ω_Q	-9.6	-2.4	-7.6	-2.7	-14.3	-4.5	-13.2	-9.1	93	89	83	91
	ω_{CL}	-3.7	-2.1	-4.9	-3.5	1.5	-1.4	-0.2	4.3	91	90	90	92
	ρ	10.6	-1.0	8.7	6.8	-14.2	3.4	-1.4	-3.0	86	95	89	94
	σ_p	0.3	-0.6	0.9	-0.6	0.3	2.6	13.0	4.3	94	94	96	95
	V_1	-0.4	-0.1	-1.3	-0.6	-6.2	0.5	-2.1	-1.8	94	94	92	92
First-order sparse	V_2	2.1	0.9	1.5	1.0	-1.0	11.5	9.5	8.8	95	95	94	94
	Q	4.1	-0.7	4.3	1.3	-6.2	124.4	131.7	102.5	92	97	93	93
	CL	0.6	0.2	-0.3	-0.0	3.0	5.8	3.0	5.1	97	96	97	97
	ω_{V_1}	-3.4	-1.1	-9.0	-1.6	8.7	10.4	15.9	10.0	97	93	86	94
	ω_{V_2}	6.7	-0.7	-13.9	2.2	-8.9	-5.4	-9.7	-8.3	94	96	86	98
	ω_{CL}	0.3	-0.8	-2.2	-1.0	3.8	7.5	6.0	8.0	94	95	92	95
	σ_p	0.4	-1.1	5.1	-1.4	3.5	-1.0	24.5	-0.8	96	96	97	93
	V_1	-1.1	-0.1	0.1	0.1	-8.9	-5.5	-8.2	-6.6	91	93	92	94
	V_2	1.1	2.6	3.4	3.0	22.7	5.1	1.7	3.5	97	91	86	91
	Q	2.9	-1.5	-2.1	-1.7	-25.8	8.3	-0.8	2.3	77	96	93	93
Mixed-order frequent	V_m	0.1	-2.3	-3.3	-2.8	38.4	-6.1	-5.6	-7.2	100	87	87	87
	K_m	-1.0	-4.9	-7.1	-6.1	-89.3	-15.2	-13.6	-17.6	19	71	75	77
	ω_{V_1}	-2.9	-2.8	-6.1	-3.7	-2.4	-6.4	-5.4	-1.1	95	91	92	94
	ω_{V_2}	-10.2	-3.6	-14.1	-5.8	-15.1	-7.6	-12.8	-9.5	93	92	81	89
	$\omega_{V_{max}}$	-2.1	-2.7	-5.0	-3.3	-2.1	-6.6	-5.1	-1.1	92	88	85	89
	σ_p	-1.1	-0.6	0.8	-0.7	-12.8	-11.4	0.3	-9.9	92	91	95	92
	Relative bias > 10% and < -10% and coverage < 0.9 are typeset in bold font												

Table 2 – Relative bias of parameter and standard error (SE) estimates and coverage rate obtained by the asymptotic method (Asym) via the Fisher information matrix (\mathbf{M}_F) and the three bootstrap methods for the unbalanced design with first-order elimination

	Relative bias of parameters (%)					Relative bias of SE (%)					Coverage rate (%)				
	Asym	B _{case,none}	B _{case,none} ^{boot}	B _{GR}	B _{GR} ^{boot}	Asym	B _{case,none}	B _{case,none} ^{boot}	B _{GR}	B _{GR} ^{boot}	Asym	B _{case,none}	B _{case,none} ^{boot}	B _{GR}	B _{GR} ^{boot}
V_1	0.1	-0.1	-0.1	-1.0	-4.6	1.2	7.3	5.2	-2.1	-2.1	97	98	97	77	98
V_2	-3.6	0.2	0.3	-1.5	3.0	-12.3	-5.4	-7.7	-0.1	1.5	91	96	92	94	96
Q	-5.7	0.5	0.5	1.3	8.1	-3.8	34.3	21.7	2.4	25.7	87	94	93	95	95
CL	-0.3	-0.1	-0.1	-0.9	-2.0	-3.4	0.1	-1.9	2.2	-1.4	95	98	97	83	96
ω_{V_1}	-2.9	-0.5	-0.6	-10.6	-7.9	-6.0	2.3	0.9	40.8	13.9	94	94	94	88	93
ω_{V_2}	10.3	1.0	1.6	6.2	19.5	8.5	-6.8	-7.9	5.1	5.5	93	96	95	81	94
ω_{CL}	-2.7	-1.5	-1.5	-6.5	-6.0	-6.8	-2.9	-4.1	20.0	5.3	91	0.93	91	87	91
σ_p	-1.2	-1.0	-0.9	5.0	6.1	-9.3	-10.3	-12.3	21.1	15.9	93	87	86	92	88

Relative bias > 10% and < -10% and coverage < 0.9 are typeset in bold font

Table 3 – Parameter estimates and their relative standard errors (RSE) obtained by the asymptotic method (Asym) and the bootstrap methods (B=999 samples) for the real dataset

	Median of parameter estimates (RSE*(%))					
	Asym	B _{case,none}	B _{case,none} ^{strat}	B _{η,GR}	B _{η,GR} ^{strat}	B _{Pη,PR}
$V_1(l)$	3.62 (1.9)	3.69 (2.2)	3.69 (2.1)	3.74 (2.1)	3.76 (2.0)	3.71 (2.0)
$V_2(l)$	2.90 (5.3)	2.89 (6.4)	2.90 (6.0)	2.96 (6.0)	2.82 (5.9)	2.88 (6.2)
$Q(l/h)$	0.14 (14.9)	0.10 (18.0)	0.10 (16.7)	0.10 (23.1)	0.08 (18.0)	0.13 (15.9)
$CL(l/h)$	0.04 (2.3)	0.04 (2.6)	0.04 (2.6)	0.04 (2.6)	0.04 (2.7)	0.04 (2.5)
$\omega_{V_1}(\%)$	20 (9.6)	21 (8.7)	21 (8.3)	17 (11.6)	16 (10.3)	19 (9.8)
$\omega_Q(\%)$	111 (12.5)	110 (13.3)	110 (13.5)	95 (22.4)	99 (16.0)	100 (9.6)
$\omega_{CL}(\%)$	29 (4.9)	29 (6.8)	29 (6.6)	30 (10.5)	29 (14.3)	29 (5.4)
$\rho_{CL,Q}$	0.90 (8.2)	0.87 (8.1)	0.86 (7.9)	0.84 (12.1)	0.88 (7.1)	0.94 (5.9)
$\sigma_p(\%)$	25 (3.8)	24 (6.2)	24 (5.8)	29 (6.7)	29 (7.4)	26 (3.5)

* RSE=SE/asymptotic parameter estimates*100

7 FIGURES

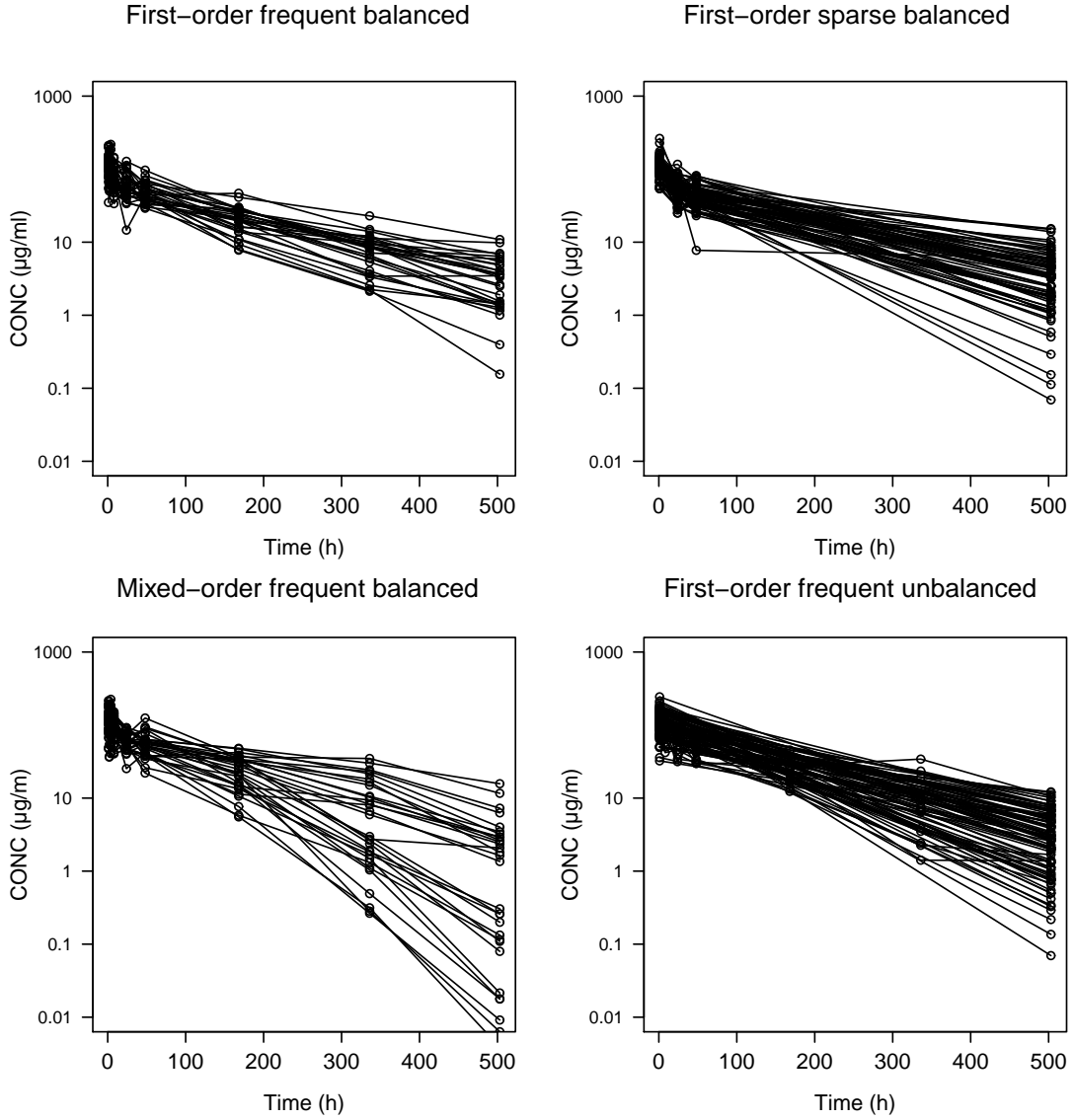


Fig. 1 – Examples of simulated data for 4 studied designs: first-order frequent sampling balanced design ($N = 30/n = 9$) (top left), first-order sparse sampling balanced design ($N = 70/n = 4$) (top right middle), mixed-order frequent sampling design ($N = 30/n = 9$) (bottom left) and first-order frequent sampling unbalanced design ($N_1 = 15/n_1 = 9$; $N_2 = 75/n_2 = 2$) (bottom right).

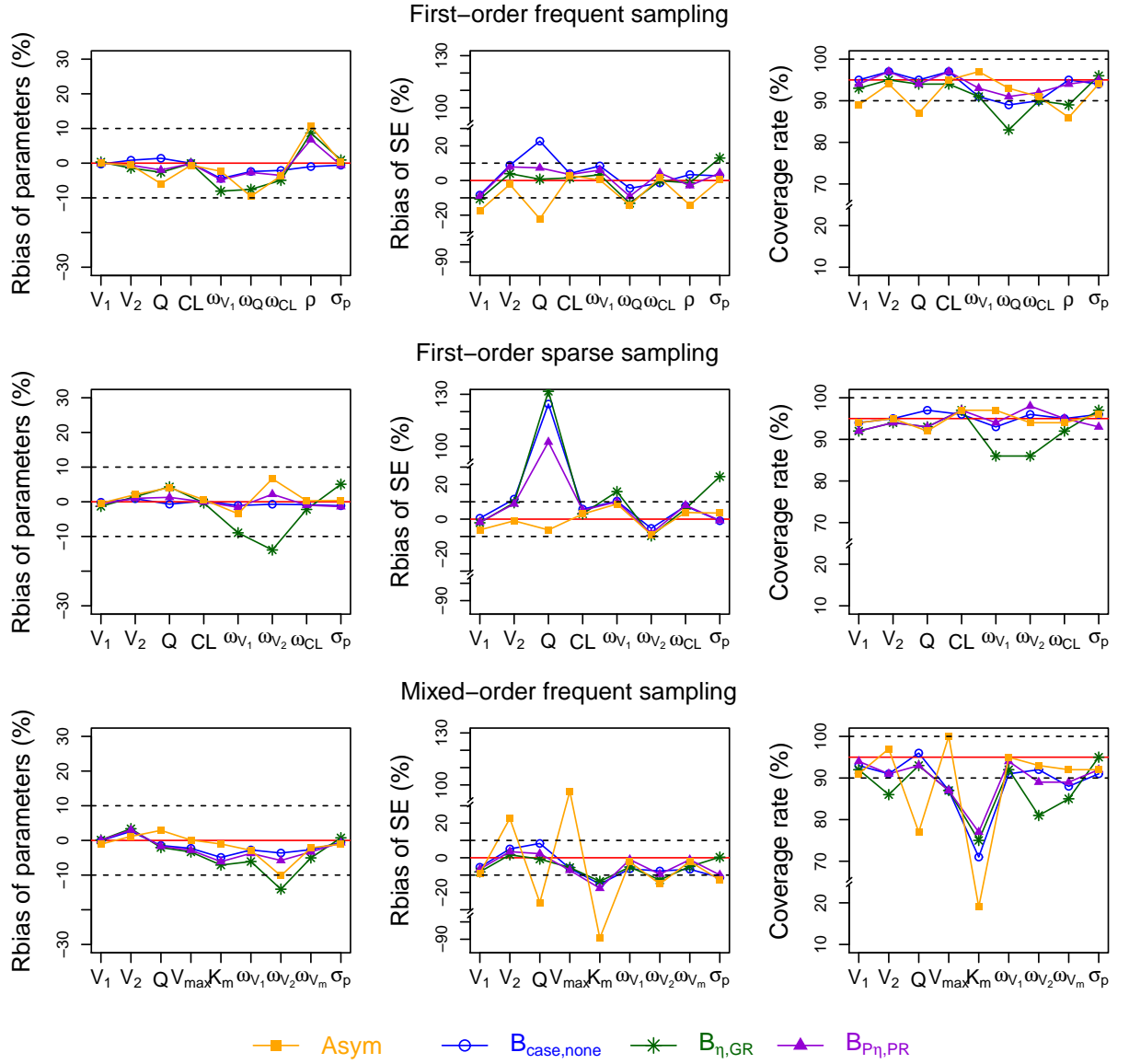


Fig. 2 – Relative bias of parameter estimates (left), relative bias of standard error (SE) estimates (middle) and coverage rate of 95% CI (right), for the asymptotic method and the three bootstrap methods in the three balanced designs: first-order frequent sampling design ($N = 30/n = 9$) (top), first-order sparse sampling design ($N = 70/n = 4$) (middle), mixed-order frequent sampling design ($N = 30/n = 9$) (bottom).

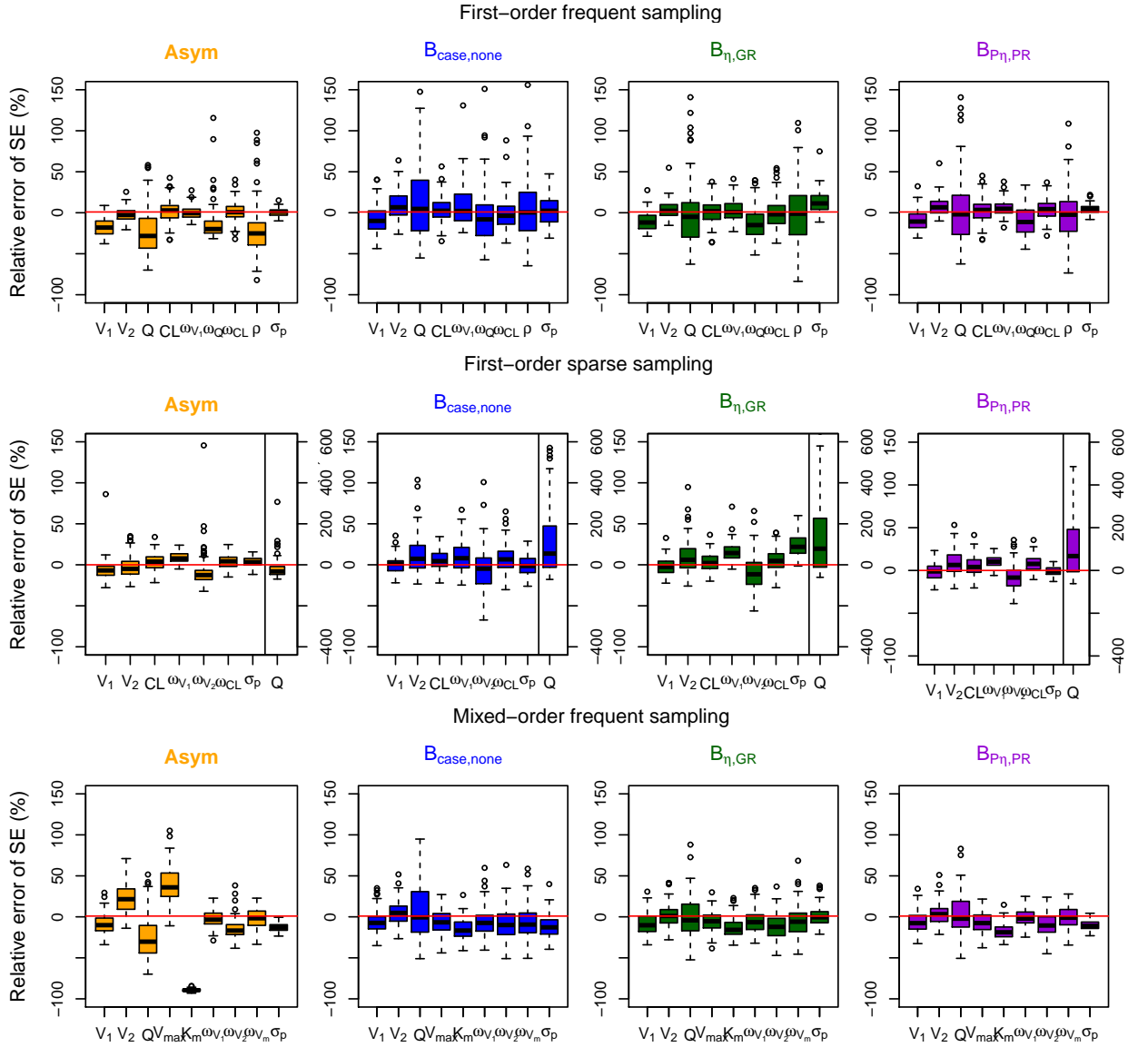


Fig. 3 – Boxplot of relative error of standard error (SE) estimates obtained by the asymptotic methods and the bootstrap methods in the three studied balanced designs: first-order frequent sampling design ($N = 30/n = 9$) (top), first-order sparse sampling design ($N = 70/n = 4$) (middle), mixed-order frequent sampling design ($N = 30/n = 9$) (bottom)

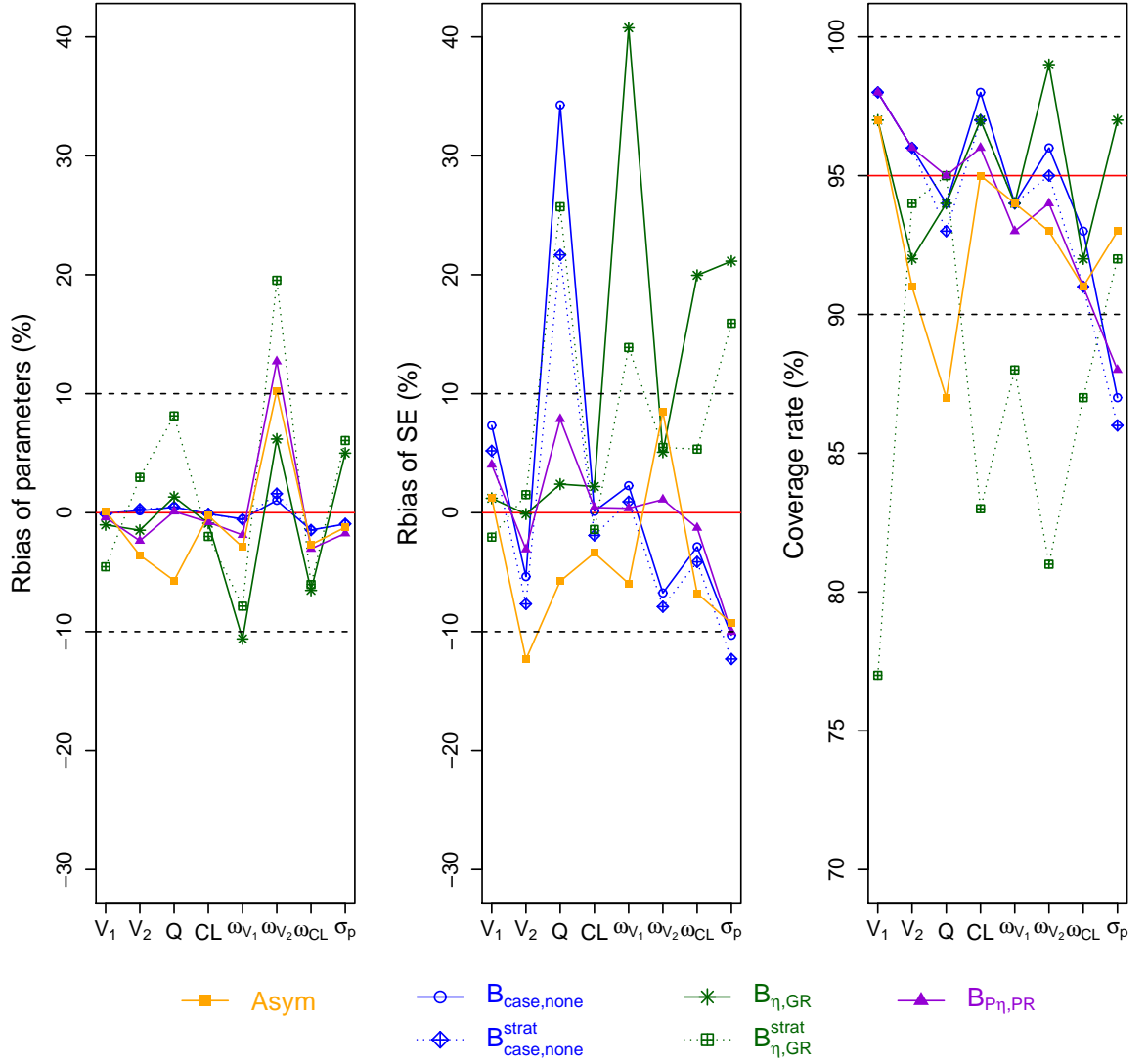


Fig. 4 – Relative bias of parameter estimates (left), relative bias of standard error (SE) estimates (middle) and coverage rate of 95% CI (right), for the asymptotic method and the three bootstrap methods in the unbalanced design ($N_1 = 15/n_1 = 9$; $N_2 = 75/n_2 = 2$) with first-order elimination

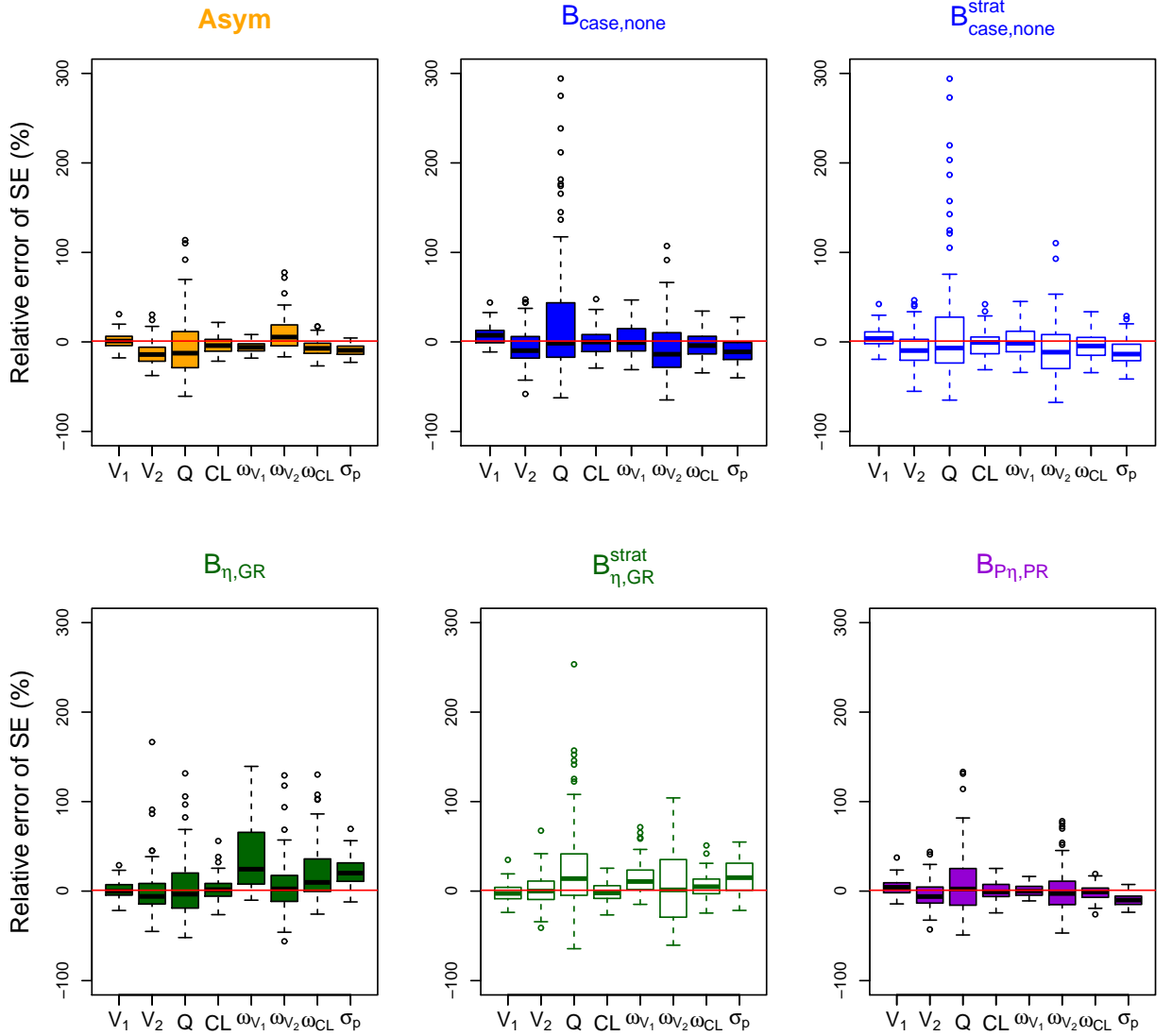


Fig. 5 – Boxplot of relative error of standard error (SE) estimates obtained by the asymptotic method and the bootstrap methods in the unbalanced design ($N_1 = 15/n_1 = 9$; $N_2 = 75/n_2 = 2$) with first-order elimination

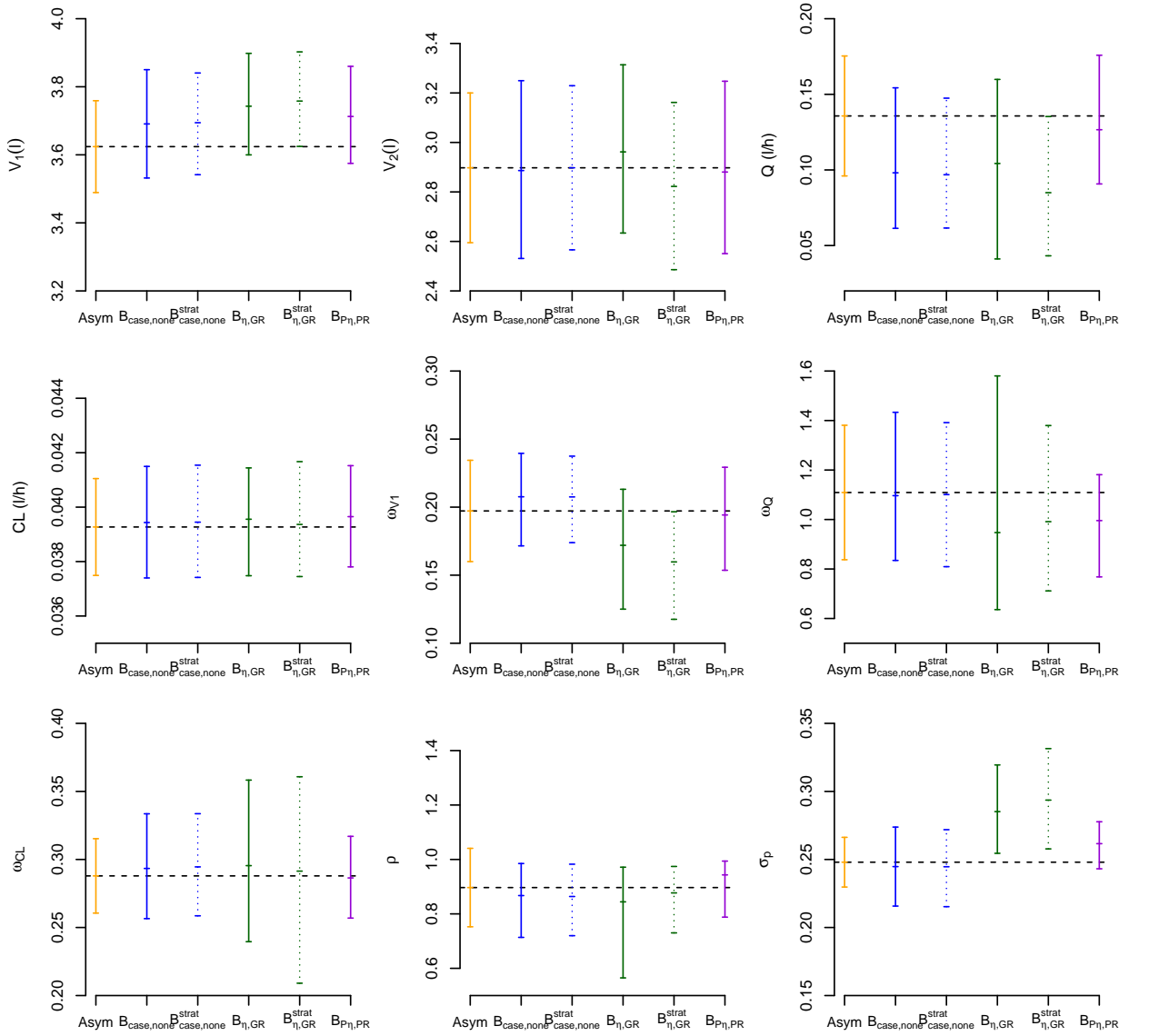


Fig. 6 – Plot of 95% confidence intervals of parameter estimates obtained by the asymptotic approach and the bootstrap methods ($B=999$ samples) for all parameters of the real dataset. The broken line in each plot represents the value of the parameter estimate obtained with the real dataset. The bootstrap confidence intervals are shown as lines in the plots.

3.2.3 Application du bootstrap aux données réelles

Suite à notre étude portant sur le comportement des approches bootstrap dans le cadre des MNLEM, nous avons illustré l'apport de l'approche bootstrap sur l'estimation de l'intertitude liée aux paramètres dans les modèles PK de l'affibercept libre et lié chez les volontaires sains (voir section 2.1) et chez les patients atteints de cancer (voir section 2.2) en utilisant le bootstrap par paires. Le bootstrap des effets aléatoires et des résidus n'a pas été appliqué à cause de la présence des données sous limite de quantification dans le modèle chez les volontaires sains et du temps de calcul sur le modèle chez les patients.

La table 3.1 présente les paramètres estimés du modèle d'affibercept libre et lié chez des volontaires sains et les résultats du bootstrap par paires avec B=999 échantillons.

TABLE 3.1 – Les paramètres du modèle TMDD de l'affibercept libre et lié chez les volontaires sains et les résultats du bootstrap

Paramètre	Données originelles		Bootstrap par paires (B=999)			Différence*(%)
	Estimation	SE	Médiane	SE	IC à 95%	
$CL(L/jour)$	0.88	0.03	0.87	0.04	0.80 - 0.94	1.2
$V_p(L)$	4.94	0.18	5.15	0.16	4.86 - 5.47	-4.2
$Q(L/jour)$	1.39	0.13	1.37	0.20	1.04 - 1.83	1.1
$V_p(L)$	2.33	0.16	2.25	0.21	1.85 - 2.66	3.5
$V_{max}(mg/jour)$	0.99	0.05	1.01	0.05	0.92 - 1.11	-2.2
$K_m(\mu g/mL)$	2.91	0.33	2.73	0.40	2.07 - 3.61	6.7
$k_{int}(1/jour)$	0.028	0.0013	0.028	0.0015	0.025 - 0.031	1.2
$\omega_{CL}(\%)$	27.98	2.79	27.57	2.94	21.52 - 32.83	1.5
$\omega_{V_p}(\%)$	27.31	2.78	28.90	5.79	17.17 - 39.89	-5.5
$\omega_Q(\%)$	49.83	7.15	46.68	8.91	29.49 - 65.32	6.8
$\omega_{V_i}(\%)$	39.77	5.51	39.25	5.77	27.25 - 50.34	1.3
$\omega_{V_{max}}(\%)$	13.57	2.28	13.32	2.54	8.54 - 18.47	1.8
$\omega_{K_m}(\%)$	45.57	6.69	47.91	7.90	31.53 - 63.09	-4.9
$\sigma_{fa}(\mu g/mL)$	0.05	0.0044	0.05	0.0078	0.03-0.06	-2.9
$\sigma_{fp}(\%)$	17.11	0.58	16.99	1.38	14.31-19.69	0.7
$\sigma_{ba}(\mu g.eq/mL)$	0.13	0.0047	0.13	0.0117	0.10-0.15	0.5

* Différence (%)=la différence relative entre l'estimation du paramètre dans les données originelles et la médiane des estimations du bootstrap

Nous avons trouvé que les paramètres estimés dans les données originelles sont très proches de ceux obtenus par le bootstrap par paires avec des différences inférieures à 6.5% et se trouvent dans les intervalles de confiance bootstrap. Par rapport aux SE des paramètres obtenues dans les données originelles, le bootstrap par paires fournit des SE comparables pour les paramètres fixes, y compris V_{max} et K_m , mais donne des SE plus grandes pour les paramètres de variabilité. La similarité dans les SE du V_{max} et du K_m obtenus par le bootstrap et avec les données originelles signifie que le modèle n'est pas très nonlinéaire.

La table 3.2 présente les paramètres du modèle d'affibercept libre et lié chez les patients et les résultats du bootstrap par paires avec B=999 échantillons.

TABLE 3.2 – Les paramètres du modèle TMDD de l'affibercept libre et lié chez les patients et les résultats du bootstrap

Paramètre	Données originelles		Bootstrap par paires (B=999)			Différence* (%)
	Estimation	SE	Médiane	SE	IC à 95%	
$CL_f(L/jour)$	0.88	0.02	0.88	0.02	0.84 - 0.91	-0.4
$\beta_{CL_f,femme}$	-0.15	0.02	-0.16	0.03	-0.21 - -0.10	-3.9
$\beta_{CL_f,poids}$	0.32	0.06	0.33	0.07	0.18 - 0.47	-0.4
$\beta_{CL_f,CLCR}$	0.18	0.04	0.19	0.04	0.10 - 0.27	-4.1
$\beta_{CL_f,ALB}$	-0.39	0.05	-0.40	0.09	-0.57 - -0.24	-2.4
$\beta_{CL_f,ALK}$	0.10	0.02	0.10	0.02	0.06 - 0.14	3.1
$\beta_{CL_f,ALT}$	-0.06	0.02	-0.06	0.02	-0.10 - -0.01	8.5
$\beta_{CL_f,gemcitabine}$	0.09	0.03	0.09	0.03	0.03 - 0.14	2.4
$V_p(L)$	4.35	0.07	4.34	0.08	4.19 - 4.49	0.1
$\beta_{V_p,femme}$	-0.21	0.02	-0.21	0.03	-0.26 - -0.16	-0.4
$\beta_{V_p,poids}$	0.39	0.06	0.40	0.06	0.28 - 0.53	-1.0
$\beta_{V_p,CLCR}$	0.10	0.04	0.10	0.04	0.02 - 0.17	-4.7
$Q(L/jour)$	1.49	0.12	1.49	0.18	1.20 - 1.90	-0.1
$V_t(L)$	3.72	0.18	3.75	0.19	3.38 - 4.17	-0.8
$V_b(L)$	4.14	0.39	4.40	0.14	4.15 - 4.70	-5.9
$V_{max}(mg/jour)$	0.82	0.08	0.86	0.02	0.82 - 0.91	-5.5
$K_m(\mu g/mL)$	1.92	0.15	1.92	0.16	1.62 - 2.25	0.1
$\beta_{K_m,femme}$	-0.23	0.08	-0.20	0.10	-0.39 - -0.01	12.6
$\beta_{K_m,ALT}$	0.26	0.12	0.22	0.15	-0.07 - 0.51	17.1
$\beta_{K_m,AST}$	-0.26	0.09	-0.19	0.12	-0.41 - 0.03	36.8
$\beta_{K_m,gemcitabine}$	0.60	0.10	0.65	0.12	0.39 - 0.87	-7.3
$CL_b(L/jour)$	0.19	0.02	0.20	0.01	0.19 - 0.21	-4.4
$\beta_{CL_b,age}$	0.18	0.05	0.18	0.05	0.07 - 0.28	2.8
$\beta_{CL_b,CLCR}$	0.09	0.03	0.09	0.03	0.03 - 0.15	-5.8
$\beta_{CL_b,ALB}$	-0.13	0.04	-0.11	0.07	-0.26 - 0.001	15.7
$\beta_{CL_b,ALT}$	-0.08	0.02	-0.08	0.03	-0.13 - -0.03	0.3
$\beta_{CL_b,irinotecan/LV/5FU}$	-0.13	0.04	-0.12	0.04	-0.20 - -0.05	8.2
$\beta_{CL_b,gemcitabine}$	0.06	0.02	0.07	0.02	0.04 - 0.11	-12.7
$\omega_{CL_f}(\%)$	30.94	0.88	31.17	1.72	27.93 - 34.63	-0.8
$\omega_{V_p}(\%)$	22.36	1.25	22.84	2.72	17.30 - 27.65	-2.1
$\omega_Q(\%)$	85.45	7.24	81.14	12.03	56.46 - 103.03	5.3
$\omega_{V_t}(\%)$	65.78	2.70	64.07	5.08	54.55 - 74.48	2.7
$\omega_{V_b}(\%)$	27.52	1.55	28.87	2.21	24.44 - 33.02	-4.7
$\omega_{K_m}(\%)$	28.22	11.45	29.96	5.58	19.91 - 42.14	-5.8
$\omega_{CL_b}(\%)$	22.30	0.78	21.59	0.83	20.01 - 23.27	3.3
$\sigma_{fa}(\mu g/mL)$	0.036	0.002	0.036	0.003	0.031 - 0.043	-0.7
$\sigma_{fp}(\%)$	32.77	0.35	32.59	0.78	31.12 - 34.13	0.6
$\sigma_{ba}(\mu g.eq/mL)$	0.34	0.01	0.34	0.04	0.255 - 0.419	0.2
$\sigma_{bp}(\%)$	9.04	0.34	9.08	1.31	6.49 - 11.74	-0.4

* Différence (%)=la différence relative entre l'estimation du paramètre dans les données originelles et la médiane des estimations du bootstrap

Nous avons également trouvé une concordance entre les paramètres obtenus par le bootstrap par paires et les paramètres estimés dans les données originelles, sauf pour les estimations de l'effet de certains covariables telles que l'effet du sexe, de ALT et de AST sur K_m et de l'effet de ALB sur CL_b . Ces différences pourraient être dues à la non stratification des covariables qui rentrent dans le modèle. Similaire aux résultats du

bootstrap sur le modèle chez les volontaires sains, les SE des paramètres de variabilité obtenues par le bootstrap sont plus grandes que celles obtenues avec les données réelles, particulièrement pour les écart-types des erreurs résiduelles.

Chapitre 4

Conclusion générale

Ce travail de thèse est consacré au développement de modèles mécanistiques pour l'affibercept (Zaltrap®), un nouveau médicament anti-angiogénique se fixant au VEGF, et utilisé dans le traitement du cancer. La construction de ces modèles repose sur l'analyse de données longitudinales en PK/PD par MNLEM. Nous nous sommes également intéressés à l'apport des approches bootstrap (EFRON et TIBSHIRANI, 1994) permettant l'estimation des incertitudes liées aux paramètres estimés. Nous avons ainsi étendu et étudié au cours de cette thèse différentes méthodes bootstrap pour les MNLEM tenant compte des caractéristiques de ces modèles (l'hétéroscédasticité, la non-linéarité, les mesures répétées au sein de chaque sujet) et les deux niveaux de variabilité (inter-sujet et résiduelle).

Les médicaments anti-angiogéniques représentent une nouvelle thérapie prometteuse pour le traitement des cancers à tumeurs solides. Ils agissent sur le microenvironnement de la tumeur en inhibant le développement des nouveaux vaisseaux sanguins indispensables à la croissance de la plupart des tumeurs solides. La majorité des anti-angiogéniques approuvés ou en développement se fixent au VEGF, un des facteurs pro-angiogéniques clés dans l'angiogenèse tumorale. Parmi ces molécules anti-VEGF, l'affibercept a une plus forte affinité *in vitro* au VEGF-A que d'autres anticorps monoclonaux, grâce à sa structure originale. Il s'agit d'une protéine de fusion comportant des domaines des récepteurs 1 et 2 du VEGF et un fragment Fc des IgG1. De plus, l'affibercept bloque non seulement le VEGF-A mais aussi le VEGF-B et le facteur de croissance placentaire (PIGF). L'activité antitumorale de l'affibercept a été observée dans les études précliniques ainsi que les études cliniques (GAYA et TSE, 2012). L'étude de la relation dose-concentration-effet de ce médicament permet de sélectionner un bon régime de dose et d'évaluer son efficacité à la fois pour l'inhibition du VEGF et la traduction de cette inhibition pour la réponse clinique. La modélisation PK/PD joue un rôle très important dans ce cas. Toutefois, de par son mode d'action sur le ligand endogène (le VEGF), la PK de ce nouveau médicament est complexe et nécessite le développement de modèles plus mécanistiques pour mieux caractériser sa distribution et sa liaison avec le VEGF.

Dans un premier temps, nous avons éclairci le mécanisme d'action de l'affibercept en développant un modèle PK de population conjoint pour l'affibercept libre et lié. Les données de concentrations plasmatiques de l'affibercept libre et lié étaient disponibles dans

plusieurs études cliniques. L'affibercept manifeste une telle affinité pour le VEGF que cette interaction se reflète dans ses propriétés PK du médicament ; ce phénomène appelé "*target-mediated drug disposition*" (TMDD) (LEVY, 1994) se caractérise par une PK non linéaire. L'étude des concentrations d'affibercept libre montre ainsi une baisse de la clairance en fonction des doses. La structure générale du modèle TMDD a été bien définie dans la littérature et a permis la description quantitative des processus biologiques sous-jacents (MAGER et JUSKO, 2001). Le modèle complet permet de prendre en compte la cinétique du médicament libre, du médicament lié ainsi que la cinétique de la cible thérapeutique. Cependant, il est complexe, souvent surparamétré à cause de la différence importante entre la vitesse du processus de liaison et celle des processus d'élimination du médicament, et de plus, son identifiabilité dépend de données généralement non mesurées, par exemple les concentrations tissulaires du ligand endogène (le VEGF), de l'affibercept libre et du lié sous forme de complexe (GIBIANSKY et GIBIANSKY, 2009). Plusieurs simplifications du modèle TMDD ont été proposées pour réduire le nombre de paramètres, par exemple remplacer la constante de vitesse d'association (k_{on}) et de dissociation (k_{off}) par la constante d'équilibre de dissociation (K_D) ou par les paramètres du Michaelis-Menten (V_{max}, K_m) en présence d'une saturation complète de la cible (GIBIANSKY et al., 2008). Le développement des modèles mécanistiques pour l'affibercept réalisé dans ce travail de thèse repose sur le modèle TMDD et ses approximations. Ce choix a été inspiré du modèle TMDD déjà développé pour le VEGF humain recombinant (rhVEGF) administré chez les patients atteints de maladie coronarienne et décrivant sa liaison avec ses récepteurs (EPPLER et al., 2002). La liaison de l'affibercept au VEGF pourrait avoir les mêmes propriétés car ce médicament comprend dans sa structure deux récepteurs du VEGF (VEGFR1 et 2).

Nous avons tout d'abord construit le modèle PK chez des volontaires sains recevant une dose unique d'affibercept (1, 2 ou 4 mg/kg) dans deux essais cliniques de phase I. Pour pouvoir identifier la structure du modèle, nous avons réalisé la construction en plusieurs étapes, d'abord le modèle de l'affibercept libre et puis le modèle conjoint de l'affibercept libre et lié. Les propriétés PK et le mécanisme d'action de l'affibercept étaient bien décrits par une approximation dite de Michaelis-Menten du modèle TMDD, comprenant deux compartiments pour l'affibercept libre, un compartiment pour l'affibercept lié et une liaison irréversible de l'affibercept au VEGF dans le compartiment périphérique. Ensuite, nous avons appliqué avec succès ce modèle aux données de patients atteints de cancer et évalué également l'influence de facteurs physiopathologiques et de la chimiothérapie concomitante sur la PK de l'affibercept. Ce modèle a permis de simuler les différents schémas d'administration et d'émettre des recommandations concernant la dose thérapeutique. Dans nos simulations, nous avons considéré les concentrations de l'affibercept lié comme un marqueur de l'inhibition du VEGF.

La liaison irréversible et de forte affinité de l'aflibercept au VEGF a lieu principalement dans le compartiment périphérique. Ce résultat a suggéré une distribution plus importante du VEGF endogène dans le compartiment périphérique par rapport au compartiment plasmatique à la fois chez les volontaires sains et chez les patients atteints de cancer. Ce phénomène a été précédemment discuté lors d'une étude de méta-analyse (KUT et al., 2007). Selon cette analyse, de grandes quantités de VEGF se trouvent dans le muscle squelettique, suggérant une source de VEGF intracellulaire. Même dans le cancer où le VEGF est libéré par la tumeur en quantité accrue, le muscle squelettique reste la source principale dans le corps (KUT et al., 2007). Par ailleurs, les taux plasmatiques de VEGF semblent être significativement élevés après l'arrêt du traitement par la plupart des thérapies anti-angiogéniques ciblant cette voie (JAIN et al., 2009). La compréhension de la distribution du VEGF dans le corps humain est particulièrement importante pour la conception et la compréhension de la thérapie anti-VEGF. Par exemple, l'administration systémique des médicaments anti-VEGF devrait prendre en compte les effets du VEGF circulant venant de son grand réservoir non dérivé de tumeurs. Il serait préférable de donner ces médicaments en continu chez les patients ayant des métastases. Comprendre la cinétique du VEGF est donc essentiel pour le pronostic et le traitement du cancer ainsi que d'autres maladies.

Grâce à ce modèle, nous avons effectué la recherche de l'effet des covariables non seulement sur les paramètres PK mais aussi sur les paramètres de liaison. Les covariables les plus importantes (le sexe, le poids corporel, l'albumine et la phosphatase alcaline) mises en évidence dans notre analyse ont également été retrouvées pour expliquer la variabilité de la PK d'un anti-VEGF similaire (le bévacizumab). Cependant, ces covariables expliquent très peu la variabilité inter-sujet des paramètres PK reflétant un impact minimal sur la cinétique de l'aflibercept et ne nécessitant pas d'adaptation posologique. Le régime de dose de 4 mg/kg toutes les deux semaines a été montré suffisant pour saturer le VEGF circulant et peut être donné à tous les patients. Le modèle développé a également permis de distinguer le cancer du pancréas par rapport à d'autres types de cancers. Bien que l'effet du type de cancer soit confondu avec l'effet de la chimiothérapie, nous attribuons cette différence à la moins grande capacité de fixation au VEGF car le cancer du pancréas dépend moins de l'angiogenèse tumorale pour survivre (LI et al., 2010). D'autres covariables telles que la charge tumorale ou les scores de gravité de maladie pourraient avoir un effet sur la PK de l'aflibercept. De plus, il serait intéressant de tester l'effet du VEGF plasmatique ou tumorale. A notre connaissance, le modèle que nous avons développé est le premier modèle PK de population basé sur le mécanisme d'action pour un médicament anti-VEGF.

Ce travail a montré le rôle très important de la méthode bioanalytique pour la modélisation PK/PD et le développement d'un nouveau médicament, en particulier les

macromolécules. La disponibilité des concentrations du médicament sous forme libre et liée a permis de caractériser à la fois l'élimination linéaire et l'élimination non linéaire, ainsi que l'internalisation du complexe (le médicament lié). Elle a également permis d'estimer les paramètres mécanistiques du système TMDD qui ne pourraient être identifiés en utilisant seulement les concentrations de l'affibercept libre. Pour les autres anti-angiogéniques, par exemple le bévacicumab et le squalamine, seule la forme totale du médicament (la somme des formes libres et liées) a généralement été dosée et leur PK a été caractérisée par un simple modèle de perfusion à deux compartiments et élimination d'ordre 1 (LU et al., 2008; BHARGAVA et al., 2001). Des techniques avancées comme la LBA (*ligand-binding assay*) avec la recherche des réactifs pertinents sont en cours de développement et de validation pour quantifier les formes libres, liées ou totales des médicaments biologiques et leurs cibles thérapeutiques (LEE et al., 2011). Ces mesures sont cruciales pour évaluer les relations dose-concentration-effet et ainsi permettre l'évaluation de l'efficacité et de la sécurité et guider la sélection de dose. Dans le cas de l'affibercept, les données du VEGF endogène circulant n'étaient pas disponibles mais les concentrations de l'affibercept lié sous forme de complexe avec le VEGF ont fournis des informations indirectes sur l'activité anti-angiogénique du produit sur la tumeur. Ainsi, la saturation de la fixation de l'affibercept au VEGF peut être observée après simulation des profils de l'affibercept lié après différentes doses, permettant ainsi la sélection de la meilleure.

Avec les résultats intéressants obtenus pour la modélisation PK, nous avons, dans un deuxième temps, développé un modèle d'inhibition de la croissance tumorale (TGI). Ce modèle prend en compte d'une part l'effet de l'affibercept sur la croissance tumorale et d'autre part, utilise les données longitudinales de la taille de la tumeur, un marqueur PD plus direct que l'affibercept lié. Il a été construit avec les données issues de l'étude VELOUR comparant l'efficacité de l'affibercept versus placebo en combinaison avec la chimiothérapie FOLFIRI (5-FU, la leucovorine et l'irinotécan) chez des patients atteints de cancer colorectal avec métastases. Nous avons étendu le modèle TGI développé par CLARET et al. (2009) prenant en compte la dynamique tumorale, l'effet antitumoral du médicament et la résistance à l'effet du médicament. Pour tenir compte du mécanisme d'action de l'affibercept et du FOLFIRI, nous avons supposé que FOLFIRI stimule la dégradation des cellules tumorales (effet sur le paramètre K_D représentant la vitesse de disparition des cellules tumorales) tandis que l'affibercept inhibe la croissance tumorale (effet sur le paramètre K_L représentant la vitesse de croissance tumorale). C'est un modèle mécanistique complexe qui combine le modèle TMDD de l'affibercept, un modèle cinétique-pharmacodynamique (KPD) du FOLFIRI qui tient compte du régime de doses administrées en l'absence de concentrations mesurées, et le modèle dynamique de la tumeur. Ce modèle TGI a décrit correctement l'évolution de la taille de la tumeur dans le groupe de l'affibercept et dans le groupe de référence. Cependant, les données de sortie d'étude n'ont été prises en compte ni modélisées dans ce modèle. Ces données sont

fréquemment observées en oncologie où les données longitudinales de biomarqueurs sont souvent associées au temps de survenue d'un évènement (IBRAHIM et al., 2010; WANG et al., 2012). Un prolongement de ce travail consisterait à les prendre en compte par modélisation conjointe de données longitudinales quantitatives et de données de survie (TSIATIS et DAVIDIAN, 2004). Cela permettra de relier l'effet sur la taille de tumeurs au bénéfice de survie chez les patients, celui-ci étant un critère d'évaluation de l'efficacité direct et plus important. Cela permettra également de prédire par simulation l'efficacité de la thérapie pour différentes stratégies de doses d'affibercept et de FOLFIRI.

Le rôle de modélisation PK/PD dans le développement d'un nouveau médicament, ici avec l'exemple de l'affibercept, a été illustré au cours de cette thèse. Grâce aux modèles mécanistiques développés, le mécanisme d'action ainsi que l'efficacité du médicament ont été mieux caractérisés en prenant en compte l'ensemble des données disponibles sur le produit depuis les phases précoces de développement (des études de phase I chez les volontaires sains) jusqu'aux phases tardives (des études de phase III chez les patients). Ce développement relie le modèle PK conjoint de l'affibercept libre et lié au modèle PD d'inhibition de la croissance tumorale et permet d'approcher du modèle de survie. Ces relations dose-concentration-effet une fois quantifiées permettent de suggérer une dose optimale afin d'atteindre un critère d'efficacité prédéfini. De plus, nous pouvons envisager d'utiliser ces modèles pour planifier des nouvelles études et s'appuyer sur l'optimisation du protocole, qui pourrait être très utile pour l'identification des paramètres dans les modèles complexes imposant d'avoir plusieurs doses différentes.

Les modèles PK/PD développés ci-dessus sont très complexes, non seulement en terme de structure mais aussi en terme de statistique, avec un certain nombre d'hypothèses sur la structure du modèle et les distributions de la variabilité. L'incertitude associée à des estimations de paramètres est généralement quantifiée par la SE obtenue par l'inverse de la matrice d'information de Fisher (MF), un calcul basée sur l'hypothèse de normalité asymptotique pour les estimateurs de paramètres et de leur SE. Cependant, les estimations d'incertitude par l'approche asymptotique peuvent être biaisées lorsque l'approximation asymptotique est incorrecte, par exemple lorsque la taille de l'échantillon est petite. Parfois, elles ne peuvent pas être obtenues à cause de problèmes d'identifiabilité du modèle complexe ou de problèmes numériques lors du calcul de l'inverse de la MF. L'approche bootstrap représente une technique alternative pour estimer la SE des paramètres, qui permet également de fournir un intervalle de confiance. Cette approche consiste à rééchantillonner les données observées avec remise pour obtenir des échantillons bootstrap de la même taille (EFRON et TIBSHIRANI, 1994). Elle est souvent utilisée dans le domaine PK/PD pour tester la stabilité de modèles en comparant les estimations asymptotiques des paramètres obtenues pour les données originales avec celles obtenues par la méthode du bootstrap et estimer leurs SE et/ou les intervalles de confiance (ETTE, 1996). Par ailleurs,

selon notre recherche bibliographique, toutes les applications PK/PD utilisent la méthode la plus simple, le bootstrap par paires, qui consiste à rééchantillonner les individus en conservant toutes leurs observations. D'autres méthodes, par exemple le bootstrap des résidus non paramétrique/paramétrique qui consiste à rééchantillonner/simuler les effets aléatoires et les résidus après un ajustement du modèle, ont été proposées afin de mieux tenir compte de la structure hiérarchique dans des données multi-niveaux, de séries temporelles ou longitudinales (VAN DER LEEDEN et al., 1997; DE LEEUW et MEIJER, 2007; HALIMI, 2005). Peu d'études ont été effectuées pour comparer ces différentes approches. Autrement dit, il n'y a pas encore de consensus sur la meilleure façon d'effectuer le bootstrap dans les MNLEM.

Nous avons voulu, au cours de cette thèse, étudier l'apport de l'approche bootstrap dans les MNLEM sur l'estimation des incertitudes liées aux paramètres estimés et donner les recommandations concernant son utilisation en pratique. Pour ce faire, nous avons proposé différentes méthodes bootstrap pour compléter celles proposées dans la littérature afin de rééchantillonner les deux niveaux de variabilité dans les données longitudinales : variabilité liée au sujet et variabilité résiduelle liée aux observations. Nous avons mis d'abord en place ces bootstraps dans le cas des MLEM, pour lequel des résultats théoriques existent et où l'approche asymptotique est satisfaisante. Des simulations basées sur des données réelles de progression naturelle de la maladie de Parkinson ont été effectuées pour comparer la performance de ces bootstraps. Bien qu'une seule variabilité interindividuelle soit rééchantillonnée, le bootstrap par paires marche très bien sur trois protocoles équilibrés simulés (riche, allégé, et avec une erreur résiduelle plus grande), ainsi que les bootstraps non paramétrique/paramétrique des effets aléatoires couplés avec le bootstrap des résidus. Pour les bootstraps des résidus non paramétriques, la correction des effets aléatoires et des résidus avant de rééchantillonner est nécessaire pour prendre en compte la sous-estimation des variances. Nous avons appliqué avec succès la correction proposée par CARPENTER et al. (2003) sur le ratio entre les matrices de variance-covariance empirique et estimée. En revanche, les méthodes de bootstrap qui rééchantillonnent seulement les résidus et le bootstrap par paires couplé avec le bootstrap des résidus avaient de mauvaises performances. En appliquant les méthodes avec de bonnes performances pour l'ensemble de données réelles, nous avons constaté des différences entre les meilleurs bootstraps, qui n'ont pas été observées lors de l'étude de simulation. Ces différences résultent de la distribution non normale des résidus et du protocole déséquilibré. Cette première simulation a permis d'identifier les meilleurs bootstraps en MLEM (le bootstrap par paires, le bootstrap non paramétrique/paramétrique des effets aléatoires et des résidus) et d'éliminer plusieurs méthodes qui ne marchent pas en MLEM.

Suite à ces résultats, nous avons évalué les trois meilleurs bootstraps dans les MNLEM, où interviennent d'autres caractéristiques telles que l'hétéroscédasticité dans le modèle

d'erreur, la non linéarité du modèle et l'hétérogénéité dans les données. Les bootstraps des effets aléatoires et des résidus ont été étendus et comparés avec le bootstrap par paires et la méthode asymptotique par simulation. Des données d'affibercept libre venant de deux études cliniques avec un modèle PK à 2 compartiments ont été utilisées pour illustrer ces simulations et l'algorithme SAEM implémenté dans le logiciel MONOLIX a été utilisé pour l'étape d'estimation. Nous avons montré que les approches bootstrap fournissent une bonne estimation de l'incertitude des estimations des paramètres dans les MNLEM. Elles sont meilleures que l'approche asymptotique dans les modèles avec une forte non linéarité. Toutefois, elles peuvent donner des estimations biaisées quand la distribution des estimations de paramètres est très asymétrique. La méthode bootstrap avec Winsorization pourrait être envisagée dans ce cas pour améliorer les estimations. (ETTE et ONYIAH, 2002). Cette méthode remplace toutes les valeurs aberrantes par un percentile prédéfini dans les données avant de calculer les statistiques. Les performances de méthodes bootstrap couplées avec la technique de Winsorization devraient être étudiées. Par ailleurs, la performance du bootstrap dépend également du nombre de réplifications (B) effectuées. Il est recommandé un B d'au moins 100 pour l'estimation de la SE et d'au moins 1000 pour l'estimation de l'intervalle de confiance (CHERNICK, 2007; BONATE, 2011). A noter que $B = 999$ est préférable par rapport à 1000 pour estimer les quantiles des intervalles de confiance sans interpolation (DAVISON et HINKLEY, 1997). Des travaux supplémentaires sur la stabilité de l'approche bootstrap et le choix optimal du B sont nécessaires pour optimiser le temps de calcul pour le bootstrap dans les MNLEM.

Nous nous attendions à ce que les bootstraps des effets aléatoires plus des résidus aient les meilleures performances dans les MNLEM, surtout dans les designs déséquilibrés car ils maintiennent la structure de données originelles. Notre simulation a confirmé la bonne performance du bootstrap paramétrique qui rééchantillonne à partir de la distribution normale, qui était la distribution utilisée pour simuler l'ensemble des données. Il serait intéressant dans une étude plus approfondie de déterminer si cette méthode reste robuste aux écarts par rapport à cette hypothèse, par exemple en simulant avec les distributions à queue lourde ou un mélange de distributions. Contrairement au bootstrap paramétrique, la performance du bootstrap des résidus non paramétrique était limitée par rapport aux autres méthodes, montrant notamment une sur-estimation des SE de variances dans le protocole déséquilibré. La correction reposant sur le ratio entre les matrices de variance-covariance empirique et estimée (le shrinkage) devrait être modifiée dans ce cas pour prendre en compte les shrinkages différents dans chaque groupe. Nous avons appliqué une stratification simple pour faire la correction dans chaque groupe mais elle a encore dégradé la performance du bootstrap. De plus, l'utilisation du bootstrap des résidus rencontre d'autres difficultés en pratique quand les données originales contiennent des données manquantes, par exemple les données sous la limite de quantification. La technique d'imputation multiple pourrait être envisagée pour remplacer les données manquantes

par les prédictions du modèle (EFRON, 1994; BROWNSTONE et VALLETTA, 2001). De même, on ne sait pas encore comment effectuer le bootstrap des résidus dans le cas de modèles à réponses multiples, tel que le modèle conjoint d'affibercept libre et lié. Les résidus peuvent être rééchantillonnés dans chaque réponse ou alors dans l'ensemble des données. Cela demande des travaux supplémentaires et c'est pourquoi nous n'avons pas appliqué cette méthode du bootstrap à la modélisation PK/PD de l'affibercept libre et lié.

Le bootstrap par paires s'est finalement avéré plus robuste que les bootstraps des résidus dans le cas de l'hétéroscédasticité, les données sous la limite de quantification et les données contenant plusieurs réponses car il rééchantillonne simplement les individus. Il est aussi plus simple, plus rapide et ne dépend d'aucune hypothèse sur le modèle. En préservant les variabilités inter-sujet et résiduelle dans une seule étape de rééchantillonnage, cette méthode fonctionne très bien dans les MLEM. Sa performance a été ensuite confirmée dans les MNLEM pour les protocoles riches équilibrés. Cependant, le bootstrap par paires ne maintient pas la même structure de données originales. C'est pourquoi il fonctionne moins bien dans les protocoles déséquilibrés malgré la stratification.

L'ensemble de nos travaux sur l'approche bootstrap dans les modèles à effets mixtes a montré que la méthode asymptotique obtient de bons résultats dans la plupart des cas, tandis que les méthodes de bootstrap fournissent de meilleures estimations de l'incertitude pour les modèles à une forte non linéarité. Des études supplémentaires sur le critère de non linéarité peuvent être envisagées pour déterminer dans quel cas la méthode asymptotique suffit et dans quel cas il vaut mieux utiliser le bootstrap en sachant que le bootstrap pour les modèles non linéaires est très lourd. Nous avons également montré que les méthodes de bootstrap peuvent être confrontées à plusieurs problèmes pratiques. Ils peuvent générer une estimation erronée de la SE d'un paramètre avec une distribution asymétrique. En outre, le rééchantillonnage dans les protocoles déséquilibrés est beaucoup plus difficile, et la stratification peut être insuffisante pour corriger l'hétérogénéité en particulier dans les protocoles très déséquilibrés.

En conclusion, ce travail de thèse a permis une meilleure compréhension du mécanisme d'action d'un nouveau anti-angiogénique (l'affibercept) en développant des modèles mécanistiques. Nous avons montré l'apport d'une démarche de modélisation et simulation PK/PD pour les médicaments anti-VEGF dans le développement de nouveaux médicaments. Elle permet de caractériser les relations dose-concentration-effet et d'optimiser la posologie. Cette démarche pourrait être utilisée pour d'autres médicaments anti-angiogéniques occupant d'autres cibles thérapeutiques en oncologie et également s'appliquer à d'autres molécules manifestant des phénomènes de type TMDD. Ce travail a également permis de confirmer la bonne performance de la méthode bootstrap par paires dans les MNLEM et suggéré d'utiliser cette méthode, en particulier pour les

modèles à une forte non linéarité, comme un outil pour l'estimation des incertitudes liés aux paramètres estimés. Toutefois, cette méthode est très coûteuse en temps de calcul et doit être utilisée de manière prudente car elle n'améliore pas nécessairement l'estimation des SE et des intervalles de confiances par rapport à l'approche asymptotique et parfois amène des performances moins bonnes.

Bibliographie

1. AARONS L., KARLSSON M.O., MENTRE F., ROMBOUT F., STEIMER J.L., VAN PEER A., Role of modelling and simulation in phase I drug development. *Eur J Pharm Sci*, 2001, vol. 13, p. 115–122
2. BAFFERT F., LE T., SENNINO B., THURSTON G., KUO C.J., HU-LOWE D., McDONALD D.M., Cellular changes in normal blood capillaries undergoing regression after inhibition of VEGF signaling. *Am J Physiol Heart Circ Physiol*, 2006, vol. 290, p. H547–559
3. BERGERS G., HANAHAN D., Modes of resistance to anti-angiogenic therapy. *Nat Rev Cancer*, 2008, vol. 8, p. 592–603
4. BHARGAVA P., MARSHALL J.L., DAHUT W., RIZVI N., TROCKY N., WILLIAMS J.I., HAIT H., SONG S., HOLROYD K.J., HAWKINS M.J., A phase I and pharmacokinetic study of squalamine, a novel antiangiogenic agent, in patients with advanced cancers. *Clin Cancer Res*, 2001, vol. 7, p. 3912–3919
5. BONATE P.L., *Pharmacokinetic-pharmacodynamic modeling and simulation*. Springer, New-York, 2 ed., 2011
6. BROWNSTONE D., VALLETTA R., The bootstrap and multiple imputations : Harnessing increased computing power for improved statistical tests. *J Econ Perspect*, 2001, vol. 15, p. 129–141
7. CARMELIET P., Mechanisms of angiogenesis and arteriogenesis. *Nat Med*, 2000, vol. 6, p. 389–395
8. CARPENTER J.R., GOLDSTEIN H., RASBASH J., A novel bootstrap procedure for assessing the relationship between class size and achievement. *Appl Statist*, 2003, vol. 52, p. 431–443
9. CHERNICK M.R., *Bootstrap methods : A Guide for Practitioners and Researchers*. John Wiley & Sons, New Jersey, 2 ed., 2007
10. CHU Q.S., Afibercept (AVE0005) : an alternative strategy for inhibiting tumour angiogenesis by vascular endothelial growth factors. *Expert Opin Biol Ther*, 2009, vol. 9, p. 263–271

11. CLARET L., GIRARD P., HOFF P.M., VAN CUTSEM E., ZUIDEVELD K.P., JORGA K., FAGERBERG J., BRUNO R., Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol*, 2009, vol. 27, p. 4103–4108
12. CONWAY E.M., COLLEN D., CARMELIET P., Molecular mechanisms of blood vessel growth. *Cardiovasc Res*, 2001, vol. 49, p. 507–521
13. COUFFINHAL T., DUFOURCQ P., DARET D., DUPLAA C., The mechanisms of angiogenesis. Medical and therapeutic applications. *Rev Med Interne*, 2001, vol. 22, p. 1064–1082
14. DAS S., KRISHEN A., Some bootstrap methods in nonlinear mixed-effect models. *J Stat Plan Inference*, 1999, vol. 75, p. 237–245
15. DAVISON A.C., HINKLEY D.V., *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, 1997
16. DE LEEUW J., MEIJER E., *Handbook of Multilevel Analysis*, chap. Resampling multilevel models. Springer, New York, 2007
17. DUDA D.G., Molecular biomarkers of response to antiangiogenic therapy for cancer. *ISRN Cell Biology*, 2012, vol. 2012, p. 11
18. EFRON B., Bootstrap methods : Another look at the jackknife. *Annal Stat*, 1979, vol. 7, p. 1–26
19. EFRON B., Missing data, imputation, and the bootstrap. *J Am Stat Assoc*, 1994, vol. 89, p. 463–475
20. EFRON B., TIBSHIRANI R.J., *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1994
21. EICHTEN A., ADLER A., COOPER B., GRIFFITH J., WEI Y., YANCOPOULOS G., LIN H., THURSTON G., Rapid decrease in tumor perfusion following vegf blockade predicts long-term tumor growth inhibition in preclinical tumor models. *Angiogenesis*, 2012, p. 1–13
22. EMEA, Draft guideline on reporting the results of population pharmacokinetic analyses. Technical report, Committee for Medicinal Products for Human Use, European Medicines Agency., 1996
23. EPPLER S.M., COMBS D.L., HENRY T.D., LOPEZ J.J., ELLIS S.G., YI J.H., ANNEX B.H., MCCLUSKEY E.R., ZIONCHECK T.F., A target-mediated model to describe the

- pharmacokinetics and hemodynamic effects of recombinant human vascular endothelial growth factor in humans. *Clin Pharmacol Ther*, 2002, vol. 72, p. 20–32
24. ETTE E.I., Comparing non-hierarchical models : application to non-linear mixed effects modeling. *Comput Biol Med*, 1996, vol. 26, p. 505–512
25. ETTE E.I., Stability and performance of a population pharmacokinetic model. *J Clin Pharmacol*, 1997, vol. 37, p. 486–495
26. ETTE E.I., ONYIAH L.C., Estimating inestimable standard errors in population pharmacokinetic studies : the bootstrap with Winsorization. *Eur J Drug Metab Pharmacokinet*, 2002, vol. 27, p. 213–224
27. FERRARA N., Vascular endothelial growth factor : basic science and clinical progress. *Endocr Rev*, 2004, vol. 25, p. 581–611
28. FERRARA N., *Angiogenesis : From Basic Ccience to Clinical Applications*. Taylor & Francis, 2007
29. FERRARA N., KERBEL R.S., Angiogenesis as a therapeutic target. *Nature*, 2005, vol. 438, p. 967–974
30. FOLKMAN J., Tumor angiogenesis : therapeutic implications. *N Engl J Med*, 1971, vol. 285, p. 1182–1186
31. FOLKMAN J., Anti-angiogenesis : new concept for therapy of solid tumors. *Ann Surg*, 1972, vol. 175, p. 409–416
32. FOOD, ADMINISTRATION D., Guidance for industry : Population pharmacokinetics. Technical report, FDA, Rockville, Maryland, USA., 1999
33. FREYER G., ISAMBERT N., YOU B., ZANETTA S., FALANDRY C., FAVIER L., TRILLET-LENOIR V., ASSADOURIAN S., SOUSSAN-LAZARD K., ZITI-LJAJIC S., FUMOLEAU P., Phase i dose-escalation study of aflibercept in combination with docetaxel and cisplatin in patients with advanced solid tumours. *Br J Cancer*, 2012, vol. 107, p. 598–603
34. GABRIELSSON J., WEINER D., *Pharmacokinetic and Pharmacodynamic Data Analysis : Concepts and Applications*. Stockholm : Apotekarsocieteten, 1999
35. GAYA A., TSE V., A preclinical and clinical review of aflibercept for the management of cancer. *Cancer Treat Rev*, 2012, vol. 38, p. 484–493
36. GE Z., BICKEL P., RICE J., An approximate likelihood approach to nonlinear mixed effects models via spline approximation. *Comput Stat Data Anal*, 2004, vol. 46, p. 747–776

37. GIBALDI M., PERRIER D., *Pharmacokinetics*. Marcel Dekker, New York, 2 ed., 1982
38. GIBIANSKY L., GIBIANSKY E., Target-mediated drug disposition model : approximations, identifiability of model parameters and applications to the population pharmacokinetic-pharmacodynamic modeling of biologics. *Expert Opin Drug Metab Toxicol*, 2009, vol. 5, p. 803–812
39. GIBIANSKY L., GIBIANSKY E., Target-mediated drug disposition model for drugs that bind to more than one target. *J Pharmacokinet Pharmacodyn*, 2010a, vol. 37, p. 323–346
40. GIBIANSKY L., GIBIANSKY E., Target-mediated drug disposition : New derivation of the michaelis-menten model, and why it is often sufficient for description of drugs with tmdd. In *19th Population Group Approach in Europe, Berlin, Gemany. Abstr 1728* [www.page-meeting.org/?abstract=1728], 2010b
41. GIBIANSKY L., GIBIANSKY E., KAKKAR T., MA P., Approximations of the target-mediated drug disposition model and identifiability of model parameters. *J Pharmacokinet Pharmacodyn*, 2008, vol. 35, p. 573–591
42. HALIMI R., *Nonlinear Mixed-effects Models and Bootstrap resampling : Analysis of Non-normal Repeated Measures in Biostatistical Practice*. VDM Verlag, Berlin, 2005
43. HALL P., Asymptotic properties of the bootstrap for heavy-tailed distributions. *Annals of Probability*, 1990, vol. 18, p. 1342–1360
44. HICKLIN D.J., ELLIS L.M., Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *J Clin Oncol*, 2005, vol. 23, p. 1011–1027
45. HILLEN F., GRIFFIOEN A.W., Tumour vascularization : sprouting angiogenesis and beyond. *Cancer Metastasis Rev*, 2007, vol. 26, p. 489–502
46. HOLASH J., DAVIS S., PAPADOPOULOS N., CROLL S.D., HO L., RUSSELL M., BOLAND P., LEIDICH R., HYLTON D., BUROVA E., IOFFE E., HUANG T., RADZIEJEWSKI C., BAILEY K., FANDL J.P., DALY T., WIEGAND S.J., YANCOPOULOS G.D., RUDGE J.S., VEGF-Trap : a VEGF blocker with potent antitumor effects. *Proc Natl Acad Sci U S A*, 2002, vol. 99, p. 11393–11398
47. HOLFORD N.H.G., CHAN P.L.S., NUTT J.G., KIEBURTZ K., SHOULSON I., PARKINSON STUDY GROUP, Disease progression and pharmacodynamics in Parkinson disease—evidence for functional protection with levodopa and other treatments. *J Pharmacokinet Pharmacodyn*, 2006, vol. 33, p. 281–311
48. IBRAHIM J.G., CHU H., CHEN L.M., Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol*, 2010, vol. 28, p. 2796–2801

49. JAIN R.K., DUDA D.G., WILLETT C.G., SAHANI D.V., ZHU A.X., LOEFFLER J.S., BATCHELOR T.T., SORENSEN A.G., Biomarkers of response and resistance to antiangiogenic therapy. *Nat Rev Clin Oncol*, 2009, vol. 6, p. 327–338
50. JUSKO W.J., KO H.C., Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clin Pharmacol Ther*, 1994, vol. 56, p. 406–419
51. KIERAN M.W., KALLURI R., CHO Y.J., The VEGF pathway in cancer and disease : responses, resistance, and the path forward. *Cold Spring Harb Perspect Med*, 2012, vol. 2, p. a006593
52. KIM J., HAYTON W.L., ROBINSON J.M., ANDERSON C.L., Kinetics of FcRn-mediated recycling of IgG and albumin in human : pathophysiology and therapeutic implications using a simplified mechanism-based model. *Clin Immunol*, 2007, vol. 122, p. 146–155
53. KUHN E., LAVIELLE M., Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal*, 2005, vol. 49, p. 1020–38
54. KUO T.T., AVESON V.G., Neonatal fc receptor and igg-based therapeutics. *MAbs*, 2011, vol. 3, p. 422–430
55. KUT C., MAC GABHANN F., POPEL A.S., Where is VEGF in the body ? a meta-analysis of VEGF distribution in cancer. *Br J Cancer*, 2007, vol. 97, p. 978–985
56. LALONDE R.L., KOWALSKI K.G., HUTMACHER M.M., EWY W., NICHOLS D.J., MILLIGAN P.A., CORRIGAN B.W., LOCKWOOD P.A., MARSHALL S.A., BENINCOSA L.J., TENSFELDT T.G., PARIVAR K., AMANTEA M., GLUE P., KOIDE H., MILLER R., Model-based drug development. *Clin Pharmacol Ther*, 2007, vol. 82, p. 21–32
57. LAVIELLE M., *MONOLIX (MOdèles NON Linéaires à effets miXtes)*. MONOLIX group, Orsay, France, 2008, <http://software.monolix.org/sdoms/software/>
58. LEE J.W., KELLEY M., KING L.E., YANG J., SALIMI-MOOSAVI H., TANG M.T., LU J.F., KAMERUD J., AHENE A., MYLER H., ROGERS C., Bioanalytical approaches to quantify "total" and "free" therapeutic antibodies and their targets : technical challenges and PK/PD applications over the course of drug development. *Aaps J*, 2011, vol. 13, p. 99–110
59. LEUNG D.W., CACHIANES G., KUANG W.J., GOEDEL D.V., FERRARA N., Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science*, 1989, vol. 246, p. 1306–1309
60. LEVY G., Pharmacologic target-mediated drug disposition. *Clin Pharmacol Ther*, 1994, vol. 56, p. 248–252

61. LI J., WIENTJES M.G., AU J.L., Pancreatic cancer : pathobiology, treatment options, and drug delivery. *Aaps J*, 2010, vol. 12, p. 223–232
62. LOGES S., SCHMIDT T., CARMELIET P., Mechanisms of resistance to anti-angiogenic therapy and development of third-generation anti-angiogenic drug candidates. *Genes Cancer*, 2010, vol. 1, p. 12–25
63. LOUIS T., Finding the observed information matrix when using the em algorithm. *J R Stat Soc Series B*, 1982, vol. 44, p. 226–233
64. LU J.F., BRUNO R., EPPLER S., NOVOTNY W., LUM B., GAUDREAULT J., Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother Pharmacol*, 2008, vol. 62, p. 779–786
65. MACKINNON J.B., Bootstrap methods in econometrics. *Econ Rec*, 2006, vol. 82, p. S2–S18
66. MAGER D.E., Target-mediated drug disposition and dynamics. *Biochem Pharmacol*, 2006, vol. 72, p. 1–10
67. MAGER D.E., JUSKO W.J., General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J Pharmacokinet Pharmacodyn*, 2001, vol. 28, p. 507–532
68. MAGER D.E., KRZYZANSKI W., Quasi-equilibrium pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *Pharm Res*, 2005, vol. 22, p. 1589–1596
69. MAGER D.E., NEUTEBOOM B., EFTHYMIPOULOS C., MUNAFO A., JUSKO W.J., Receptor-mediated pharmacokinetics and pharmacodynamics of interferon-beta1a in monkeys. *J Pharmacol Exp Ther*, 2003, vol. 306, p. 262–270
70. PARKE J., HOLFORD N.H., CHARLES B.G., A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models. *Comput Methods Programs Biomed*, 1999, vol. 59, p. 19 – 29, ISSN 0169-2607
71. PINHEIRO J.C., BATES D.M., *Mixed-effects models for S and S-plus*. Springer, New York, 2000
72. R DEVELOPMENT CORE TEAM, *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008
73. RAMLAU R., GORBUNOVA V., CIULEANU T.E., NOVELLO S., OZGUROGLU M., GOKSEL T., BALDOTTO C., BENNOUNA J., SHEPHERD F.A., LE-GUENNEC S., REY A., MILLER V., THATCHER N., SCAGLIOTTI G., Afibercept and docetaxel versus

- docetaxel alone after platinum failure in patients with advanced or metastatic non-small-cell lung cancer : a randomized, controlled phase III trial. *J Clin Oncol*, 2012, vol. 30, p. 3640–3647
74. RAPISARDA A., MELILLO G., Overcoming disappointing results with antiangiogenic therapy by targeting hypoxia. *Nat Rev Clin Oncol*, 2012, vol. 9, p. 378–390
75. RIBBA B., KALOSHI G., PEYRE M., RICARD D., CALVEZ V., TOD M., CAJAVEC-BERNARD B., IDBAIH A., PSIMARAS D., DAINESE L., PALLUD J., CARTALAT-CAREL S., DELATTRE J.Y., HONNORAT J., GRENIER E., DUCRAY F., A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clin Cancer Res*, 2012, vol. 18, p. 5071–5080
76. SAMSON A., LAVIELLE M., MENTRE F., Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model : Application to hiv dynamics model. *Comput Stat Data An*, 2006, vol. 51, p. 1562–1574
77. SATCHI-FAINARO R., DUNCAN R., BARNES C., Polymer therapeutics for cancer : current status and future challenges. *Adv Polym Sci*, 2006, vol. 193, p. 1–65
78. SEGRAVE A.M., MAGER D.E., CHARMAN S.A., EDWARDS G.A., PORTER C.J., Pharmacokinetics of recombinant human leukemia inhibitory factor in sheep. *J Pharmacol Exp Ther*, 2004, vol. 309, p. 1085–1092
79. SHAO J., TU D., *The Jackknife and Bootstrap*. Springer, New York, 1995
80. SHEINER L., BEAL S., *NONMEM Version 51*. University of California, NONMEM Project Group, San Francisco, 1998
81. SOPHIE R., AKHTAR A., SEPAH Y., IBRAHIM M., BITTENCOURT M., DO D., NGUYEN Q., Aflibercept : a potent vascular endothelial growth factor antagonist for neovascular age-related macular degeneration and other retinal vascular diseases. *Biol Ther*, 2012, vol. 2, p. 1–22
82. THAI H.T., MENTRE F., HOLFORD N.H., VEYRAT-FOLLET C., COMETS E., A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharm Stat*, 2013, doi :10.1002/pst.1561
83. THAI H.T., VEYRAT-FOLLET C., VIVIER N., DUBRUC C., SANDERINK G., MENTRE F., COMETS E., A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects. *Br J Clin Pharmacol*, 2011, vol. 72, p. 402–414

84. THAM L.S., WANG L., SOO R.A., LEE S.C., LEE H.S., YONG W.P., GOH B.C., HOLFORD N.H., A pharmacodynamic model for the time course of tumor shrinkage by gemcitabine + carboplatin in non-small cell lung cancer patients. *Clin Cancer Res*, 2008, vol. 14, p. 4213–4218
85. TSIATIS A.A., DAVIDIAN M., Joint modeling of longitudinal and time-to-event data : an overview. *Stat Sinica*, 2004, vol. 14, p. 809–834
86. VAN CUTSEM E., TABERNERO J., LAKOMY R., PRENEN H., PRAUSOVA J., MACARULLA T., RUFF P., VAN HAZEL G.A., MOISEYENKO V., FERRY D., MCKENDRICK J., POLIKOFF J., TELLIER A., CASTAN R., ALLEGRA C., Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase iii randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J Clin Oncol*, 2012, vol. 30, p. 3499–506
87. VAN DER LEEDEN R., BUSING F.M.T.A., MEIJER E., Bootstrap methods for two-level models. Technical Report PRM 97-04, Leiden University, Department of Psychology, Leiden, 1997
88. VONESH E.F., A note on the use of laplace’s approximation for nonlinear mixed-effects models. *Biometrika*, 1996, vol. 83, p. 447–452
89. WANG P., SHEN W., BOYE M.E., Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. *Health Serv Outcomes Res Methodol*, 2012, vol. 12, p. 182–199
90. WANG Y., SUNG C., DARTOIS C., RAMCHANDANI R., BOOTH B.P., ROCK E., GOBBURU J., Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin Pharmacol Ther*, 2009, vol. 86, p. 167–174
91. WEHRENS R., PUTTER H., BUYDENS L.M.C., The bootstrap : a tutorial. *Chemom Intell Lab Syst*, 2000, vol. 54, p. 35–52
92. WOO S., KRZYZANSKI W., JUSKO W.J., Target-mediated pharmacokinetic and pharmacodynamic model of recombinant human erythropoietin (rHuEPO). *J Pharmacokinet Pharmacodyn*, 2007, vol. 34, p. 849–868
93. WU H., ZHANG J.T., The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Stat Med*, 2002, vol. 21, p. 3655–3675
94. YAFUNE A., ISHIGURO M., Bootstrap approach for constructing confidence intervals for population pharmacokinetic parameters. I : A use of bootstrap standard error. *Stat Med*, 1999, vol. 18, p. 581–599

95. YANCOPOULOS G.D., DAVIS S., GALE N.W., RUDGE J.S., WIEGAND S.J., HOLASH J.,
Vascular-specific growth factors and blood vessel formation. *Nature*, 2000, vol. 407, p.
242–248

Résumé

L'angiogenèse, la croissance de nouveaux vaisseaux sanguins à partir de vaisseaux préexistants, joue un rôle crucial dans la croissance des tumeurs malignes et le développement des métastases. Elle est médiée notamment par le facteur de croissance vasculaire endothélial (VEGF), cible thérapeutique de nouveaux médicaments anti-angiogéniques comme l'affibercept (Zaltrap[®], développé conjointement par Regeneron et Sanofi). Il s'agit d'une protéine de fusion comportant des domaines des récepteurs VEGFR-1 et VEGFR-2 et un fragment Fc des IgG1. Il bloque le VEGF A, le VEGF-B ainsi que le facteur de croissance placentaire (PlGF) et donc l'angiogenèse. Du fait de cette liaison, les propriétés pharmacocinétique (PK)/pharmacodynamique (PD) de ce nouveau médicament deviennent plus complexes.

Dans cette thèse, nous avons étudié le mécanisme d'action de l'affibercept en développant des modèles PK/PD de population. Nous avons tout d'abord construit le modèle PK conjoint de l'affibercept libre et lié chez les volontaires sains grâce aux données riches. Nous avons ensuite appliqué avec succès ce modèle aux données chez les patients atteints de cancer et étudié également l'influence de facteurs physiopathologiques sur leur PK. Ce modèle a permis de simuler les différents schémas d'administration et de supporter le choix de dose thérapeutique. Afin de mieux évaluer l'efficacité de l'affibercept, nous avons par la suite construit un modèle PD caractérisant l'inhibition de la croissance tumorale sous l'effet combiné de l'affibercept et du FOLFIRI (5-fluorouracile, la leucovorine et l'irinotécan) chez les patients atteints du cancer colorectal métastatique.

L'incertitude liée à l'estimation des paramètres dans des modèles complexes peut être biaisée et parfois n'est pas obtenue. Nous avons donc étudié par simulation l'incertitude des paramètres obtenue par différentes méthodes de bootstrap permettant de rééchantillonner deux niveaux de variabilité (inter-sujet et résiduelle) dans les modèles non linéaires à effets mixtes (MNLEM). Ainsi, nous avons montré que le bootstrap ne fournit de meilleures estimations de l'incertitude des paramètres que dans les MNLEM avec une forte non linéarité par rapport à l'approche asymptotique. Le bootstrap par paires fonctionne aussi bien que le bootstrap non paramétrique des effets aléatoires et des résidus. Cependant, ils peuvent être confrontés à des problèmes pratiques, par exemple des distributions asymétriques dans les estimations des paramètres et des protocoles déséquilibrés où la stratification pourrait être insuffisante.

Abstract

Angiogenesis, the development of new blood vessels from pre-existing vasculature, plays a crucial role in the growth of malignant tumors and the development of metastases. It is particularly mediated by vascular endothelial growth factor (VEGF), a therapeutic target of new anti-angiogenic drugs such as aflibercept (Zaltrap[®], developed by Regeneron and Sanofi). It is a recombinant fusion protein consisting of human VEGF receptor extracellular domains (VEGFR1 and VEGFR2) fused to the Fc portion of human IgG1. It binds to VEGF-A, VEGF-B as well as to placental growth factor (PlGF). Because of this binding, the pharmacokinetic (PK)/pharmacodynamic (PD) properties of this new drug become more complex.

In this thesis, we have studied the mechanism of action of aflibercept by building population PK/PD models. We firstly developed the joint PK model of free and bound aflibercept in healthy subjects thanks to rich data. We then applied this model to data in cancer patients and also assessed the influence of physiopathologic factors on their PK. This model allowed to simulate different dosing regimens and to support the choice of therapeutic dose. To better evaluate the efficacy of aflibercept, we also build a PD model for tumor growth inhibition under the effect of aflibercept and FOLFIRI (5-fluorouracil, leucovorin and irinotecan) in patients with metastatic colorectal cancer.

The uncertainty in parameter estimates in complex models may be biased or sometimes cannot be obtained. We therefore investigated by simulation the uncertainty obtained by different bootstrap methods allowing resampling two levels of variability (between subject and residual) in nonlinear mixed-effects models (NLMEM). We have shown that the bootstraps only provide better estimates of uncertainty in NLMEM with high nonlinearity compared to the asymptotic method. The case bootstrap performs as well as the nonparametric bootstrap of both random effects and residuals. However, they may face practical problems, e.g skewed distributions in parameter estimates and unbalanced designs where stratification may be insufficient.

Mots clés : Pharmacocinétique ; Pharmacodynamique ; Anti-angiogénique ; Modèles non linéaires à effets mixtes ; Bootstrap ; Incertitude des paramètres